

UNIVERSIDADE FEDERAL DO PARANÁ

MARIA CAROLINA MARTINS PEREIRA

**EMPREGO DE REDES NEURAIS ARTIFICIAIS RBF NA PREVISÃO DE
NEVOEIRO NO AEROPORTO INTERNACIONAL AFONSO PENA**

CURITIBA

2016

MARIA CAROLINA MARTINS PEREIRA

**EMPREGO DE REDES NEURAIS ARTIFICIAIS RBF NA PREVISÃO DE
NEVOEIRO NO AEROPORTO INTERNACIONAL AFONSO PENA**

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre em Métodos
Numéricos em Engenharia, no Programa de Pós-
graduação em Métodos Numéricos em
Engenharia, do Setor de Tecnologia, da
Universidade Federal do Paraná.
Orientador: Prof. Dr. Ricardo C. de Almeida

CURITIBA

2016

Pereira, Maria Carolina Martins

Emprego de redes neurais artificiais RBF na previsão de nevoeiro
no aeroporto Afonso pena / Maria Carolina Martins Pereira. –
Curitiba, 2016.
85 f. : il.

Dissertação (mestrado) – Universidade Federal do Paraná, Setor de
Tecnologia, Programa de Pós Graduação em Métodos Numéricos em
Engenharia

Orientador: Ricardo C. de Almeida

Bibliografia: p. 80-82

1. Redes neurais (Computação). 2. Aeroportos - Visibilidade. I.
Almeida, Ricardo C. de. II. Título.

CDD 624.171



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor Tecnologia
Programa de Pós Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA
Código CAPES: 40001016030P0

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **MARIA CAROLINA MARTINS PEREIRA**, intitulada: **"EMPREGO DE REDES NEURAIS ARTIFICIAIS RBF NA PREVISÃO DE NEVOEIRO NO AEROPORTO INTERNACIONAL AFONSO PENA."**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua aprovação.

Curitiba, 31 de Agosto de 2016.

Prof RICARDO CARVALHO DE ALMEIDA
Presidente da Banca Examinadora (UFPR)

Prof EMÍLIO GRACILIANO FERREIRA MERCURI
Avaliador Interno (UFPR)

Prof REINALDO BÔMFIM DA SILVEIRA
Avaliador Externo (SIMEPAR)

Dedico este trabalho à minha mãe:
rocha firme e abrigo seguro.

AGRADECIMENTOS

A Deus, por permitir mais esta conquista.

À minha mãe, pela motivação e apoio incondicionais.

Ao Prof. Dr. Ricardo Carvalho de Almeida, pela generosidade em compartilhar seus conhecimentos.

À Força Aérea Brasileira, pela concessão dos dados que permitiram esta pesquisa.

“Nem olhos viram, nem ouvidos ouviram, nem jamais penetrou em coração humano o que Deus tem preparado para aqueles que O amam.” (1Co 2.9)

RESUMO

Restrições de visibilidade severas causadas por nevoeiros são um grande empecilho à aviação, pois podem causar atrasos ou cancelamentos de voos, além de trazerem grandes riscos às operações aéreas. O Aeroporto Internacional Afonso Pena, localizado na cidade de São José dos Pinhais, no estado do Paraná, é o aeródromo brasileiro que mais sofre com estes fenômenos meteorológicos, passando, em média 400 horas por ano impedido de realizar pousos e decolagens.

O fenômeno meteorológico nevoeiro caracteriza-se como uma formação densa de gotículas de água, restringindo a visibilidade horizontal a menos de 1000 metros. A previsão do fenômeno nevoeiro desafia meteorologistas ao redor do mundo, pois sua formação, duração e extinção estão entre as previsões meteorológicas mais difíceis de serem realizadas com eficácia.

Redes Neurais Artificiais (RNA) são ferramentas computacionais inspiradas no funcionamento do sistema nervoso humano, capazes de identificar padrões e aprender através da experiência. Das diversas configurações existentes, as RNA de Função de Base Radial são especialmente capazes de realizar classificações. Esta configuração de RNA foi adotada para este trabalho por apresentar esta característica.

Esta pesquisa objetiva verificar a eficácia do emprego de RNAs na previsão categórica do fenômeno nevoeiro para até três horas, bem como detectar quais as variáveis meteorológicas influenciam mais significativamente na eficácia da previsão. Para tanto, foram utilizadas três RNAs, uma para cada hora de previsão, treinadas independentemente a partir de dados meteorológicos coletados na Estação Meteorológica de Superfície e na Estação Meteorológica de Altitude localizadas no Aeroporto Afonso Pena. A amostra de dados completa compreende o período dos anos de 1997 a 2014.

A partir destes dados foi possível realizar uma análise sobre a persistência do fenômeno, bem como realizar estudo climatológico sobre o comportamento das diversas variáveis meteorológicas quando na ocorrência do nevoeiro, a fim de compreender melhor a gênese do evento, e também no

intuito de estabelecer um modelo conceitual que auxilie na sua previsão. Além disso, o desempenho destas técnicas de previsão foi comparado ao desempenho da técnica de previsão através de redes neurais, a fim de averiguar-se a eficácia desta última.

As RNAs utilizadas para a previsão do fenômeno passaram por diversas fases de treinamento, até se chegar às configurações ótimas, cujos dados de entrada são séries temporais com quatro horas, apenas de dados de visibilidade, e a camada intermediária dispõe de dois neurônios ocultos.

A técnica de previsão através de climatologia apresentou razão de falso alarme em torno de 96%, superestimando a ocorrência do fenômeno, aproximadamente, seis vezes; a técnica de previsão através de persistência apresentou índices de desempenho bastante próximos aos apresentados pelas redes neurais; no entanto, a técnica de previsão de nevoeiros através do uso de RNAs alcançou a maior probabilidade de detecção, indicando o desempenho superior desta técnica na previsão do fenômeno nevoeiro para o Aeroporto Afonso Pena.

ABSTRACT

Severe visibility restrictions caused by fog are a major impediment to aviation since they may cause flight delays or cancellations, in addition to bringing serious risks to air operations. The Afonso Pena International Airport, located in the city of São José dos Pinhais, in the state of Paraná, is the Brazilian airport most affected by this weather phenomenon, being, on average, 405,5 hours per year prevented from performing landings and takeoffs.

The meteorological phenomenon fog is characterized by a dense formation of water droplets which restricts horizontal visibility to less than 1.000 meters. Fog forecasts challenge meteorologists around the world and its duration and termination are among the most difficult weather forecasts to be effectively made.

Artificial Neural Networks (ANN) are computer tools inspired by the functioning of the human nervous system that are able to identify patterns and learn from experience. Among the various existing configurations, the Radial Basis Functions ANNs are especially suitable to perform pattern classification. Such a type of ANN was used in this work due to that characteristic.

The major goal of this research is to verify the effectiveness of the use of ANNs in categorical forecast of fog for up to three hours ahead, as well as to detect which meteorological variables have more significant influence on its formation process. For that, three ANNs were used, one for each forecast hour. Each of these ANNs were independently trained with meteorological data collected from the Surface and Upper Air Meteorological Stations located at the Afonso Pena Airport. The full data sample covers the period between 1997 and 2014.

From these data, it was possible to perform an analysis of the persistence of the phenomenon as well as a climatological study on the behavior of the several meteorological variables involved in fog occurrence. Those analyses allowed a better understanding of the genesis of the event and the establishment of a conceptual model that helps to forecast the phenomenon. In addition, the performance of such forecast techniques was

compared to the performance of the forecast technique through neural networks in order to investigate the effectiveness of the latter.

The ANNs used for the forecasts went through several training stages until reaching the optimal configurations, whose input data are four-hour time series of visibility data only, and the intermediate layer has two hidden neurons.

The forecast technique through climatology had a false-alarm rate of 96%, overestimating the occurrence of the phenomenon about six times; the forecast technique through persistence had performance indices very close to those showed by the neural networks; however, the fog forecast technique with the use of ANNs reached the highest detection probability, indicating the superior performance of this technique in the forecast of the fog phenomenon for the Afonso Pena Airport.

LISTA DE FIGURAS

FIGURA 1-	REPRESENTAÇÃO DE UM NEURÔNIO BIOLÓGICO	6
FIGURA 2-	REPRESENTAÇÃO UM NEURÔNIO ARTIFICIAL	8
FIGURA 3-	FUNÇÃO GAUSSIANA	12
FIGURA 4-	REPRESENTAÇÃO DA OPERAÇÃO DE UMA RNA RBF	12
FIGURA 5-	FLUXOGRAMA DA PRIMEIRA FASE DO TREINAMENTO	15
FIGURA 6-	FLUXOGRAMA DA SEGUNDA FASE DO TREINAMENTO	17
FIGURA 7-	OROGRAFIA DAS VIZINHAS DA CIDADE DE SÃO JOSÉ DOS PINHAIS	23
FIGURA 8-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NOS MESES DO ANO	37
FIGURA 9-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NAS HORAS DO DIA	38
FIGURA 10-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A TEMPERATURA DO AR	39
FIGURA 11-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A UMIDADE RELATIVA	41
FIGURA 12-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A COBERTURA DE NUVENS	42
FIGURA 13-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRESSÃO ATMOSFÉRICA	43
FIGURA 14-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO EM QUADRANTES	45
FIGURA 15-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A VELOCIDADE DO VENTO	47
FIGURA 16-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRECIPITAÇÃO ACUMULADA	48
FIGURA 17-	EVOLUÇÃO TEMPORAL DA CORRELAÇÃO ENTRE AS VARIÁVEIS COMPONENTES DA AMOSTRA E A	

	VARIÁVEL VISIBILIDADE	54
FIGURA 18-	CONVERGÊNCIA DO ERRO DA REDE R1 NA PRIMEIRA FASE DE TREINAMENTO	59
FIGURA 19-	CONVERGÊNCIA DO ERRO DA REDE R2 NA PRIMEIRA FASE DE TREINAMENTO	60
FIGURA 20-	CONVERGÊNCIA DO ERRO DA REDE R3 NA PRIMEIRA FASE DE TREINAMENTO	60
FIGURA 21-	CONVERGÊNCIA DO ERRO DA REDE R1 NA TERCEIRA FASE DE TREINAMENTO	64
FIGURA 22-	CONVERGÊNCIA DO ERRO DA REDE R2 NA TERCEIRA FASE DE TREINAMENTO	65
FIGURA 23-	CONVERGÊNCIA DO ERRO DA REDE R3 NA TERCEIRA FASE DE TREINAMENTO	65
FIGURA 24-	CONVERGÊNCIA DO ERRO DA REDE R1 NA QUINTA FASE DE TREINAMENTO	71
FIGURA 25-	CONVERGÊNCIA DO ERRO DA REDE R2 NA QUINTA FASE DE TREINAMENTO	72
FIGURA 26-	CONVERGÊNCIA DO ERRO DA REDE R3 NA QUINTA FASE DE TREINAMENTO	72
FIGURA 27-	CONVERGÊNCIA DO ERRO DA REDE R1 NA SEXTA FASE DE TREINAMENTO	74
FIGURA 28-	CONVERGÊNCIA DO ERRO DA REDE R2 NA SEXTA FASE DE TREINAMENTO	74
FIGURA 29-	CONVERGÊNCIA DO ERRO DA REDE R3 NA SEXTA FASE DE TREINAMENTO	75

LISTA DE TABELAS

TABELA 1-	ESQUEMA DOS PROCEDIMENTOS COMPUTACIONAIS PARA A PRIMEIRA FASE DO TREINAMENTO	14
TABELA 2-	ESQUEMA DOS PROCEDIMENTOS COMPUTACIONAIS PARA A SEGUNDA FASE DO TREINAMENTO.....	17
TABELA 3-	POLARIZAÇÃO DA VARIÁVEL DIREÇÃO DO VENTO	29
TABELA 4-	NÚMERO DE REGISTROS DA AMOSTRA DE DADOS ..	31
TABELA 5-	TABELA DE CONTINGÊNCIA	32
TABELA 6-	HIERARQUIA ENTRE AS MÉTRICAS	35
TABELA 7-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NOS MESES DO ANO	36
TABELA 8-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NAS HORAS DO DIA	37
TABELA 9-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A TEMPERATURA DO AR	39
TABELA 10-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A UMIDADE RELATIVA	40
TABELA 11-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A COBERTURA DE NUVENS	42
TABELA 12-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRESSÃO ATMOSFÉRICA	43
TABELA 13-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO	44
TABELA 14-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO EM QUADRANTES	45
TABELA 15-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A VELOCIDADE DO VENTO	46
TABELA 16-	DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRECIPITAÇÃO ACUMULADA	47

TABELA 17-	CONDIÇÕES FAVORÁVEIS À FORMAÇÃO DE NEVOEIRO	49
TABELA 18-	EXEMPLO DE PREVISÃO POR PERSISTÊNCIA	51
TABELA 19-	DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIROS POR CLIMATOLOGIA	55
TABELA 20-	DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIROS POR PERSISTÊNCIA	57
TABELA 21-	RESULTADOS DOS EXPERIMENTOS COM SÉRIES TEMPORAIS CONSTITUÍDAS DE DADOS DE VISIBILIDADE NO FORMATO BIPOLAR	59
TABELA 22-	MELHOR DESEMPENHO DA PRIMEIRA ETAPA DE TREINAMENTO	61
TABELA 23-	PRIMEIRO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS	62
TABELA 24-	MELHOR DESEMPENHO DA SEGUNDA ETAPA DE TREINAMENTO	63
TABELA 25-	PRIMEIRA INSERÇÃO DE VARIÁVEIS	63
TABELA 26-	MELHOR DESEMPENHO DA TERCEIRA ETAPA DE TREINAMENTO	66
TABELA 27-	SEGUNDO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS	66
TABELA 28-	MELHOR DESEMPENHO NO SEGUNDO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS	67
TABELA 29-	MELHOR DESEMPENHO DA QUARTA ETAPA DE TREINAMENTO	68
TABELA 30-	DISTRIBUIÇÃO DOS VALORES DE VISIBILIDADE	69
TABELA 31-	DESEMPENHO DAS REDES NEURAIS COM AMOSTRAS AJUSTADAS	70
TABELA 32-	MELHOR DESEMPENHO DA QUINTA ETAPA DE TREINAMENTO	72
TABELA 33-	TERCEIRO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS	73
TABELA 34-	MELHOR DESEMPENHO DA SEXTA ETAPA DE	

	TREINAMENTO	75
TABELA 35-	DESEMPENHO DA OPERAÇÃO DAS RNA	76
TABELA 36-	COMPARATIVO ENTRE OS RESULTADOS OBTIDOS ..	77

LISTA DE SÍMBOLOS

-3h-	Três horas antes da ocorrência
-2h-	Duas horas antes da ocorrência
-1-	Não classe
-1 -1 -1 1-	Direção do vento no quarto quadrante
-1 -1 1 -1-	Direção do vento no terceiro quadrante
-1 1 -1 -1-	Direção do vento no segundo quadrante
-1h-	Uma hora antes da ocorrência
+1h-	Uma hora subsequente
+2h-	Duas horas subsequentes
+3h-	Três horas subsequentes
δ_k -	Produto da diferença entre os elementos de referência e os de saída, e o sinal produzido pela derivada da k-ésima função de ativação da camada de saída
ϵ -	Vetor saída de referência
θ -	Limiar de ativação
μ_σ -	Passos de treinamento das variâncias
μ_c -	Passos de treinamento das posições centrais
μ_W -	Passos de treinamento dos pesos sinápticos
φ_j -	Sinal produzido pelo j-ésimo neurônio oculto
Σ -	Combinador linear
σ -	Desvio padrão da gaussiana (<i>cluster</i>)
ϕ -	Função de base radial
0h-	Momento da ocorrência
1-	Classe
1 -1 -1 -1-	Direção do vento no primeiro quadrante
a-	Número de casos previstos e observados
b-	Número de casos previstos e não observados
c-	Número de casos não previstos e observados
c_j -	Representante central do j-ésimo neurônio oculto
d-	Número de casos não previstos e não observados
E-	Erro do vetor de saída

f_k -	Funções de ativação da camada de saída
g_j -	Funções de ativação da camada oculta
n -	Número de iterações ou casos
p_i -	Valores de variáveis meteorológicas
r -	Correlação
t -	Vetor posição do centro da gaussiana (<i>cluster</i>)
u -	Potencial de ativação
v_i -	Valores de visibilidade
W_i -	Pesos sinápticos
$W_{ij}^{(1)}$ -	Pesos sinápticos da ligação entre o neurônio i e o neurônio oculto j
	Pesos sinápticos da ligação entre o neurônio oculto j e o neurônio de saída k
$W_{jk}^{(2)}$ -	Somatório dos sinais de entrada do k -ésimo neurônio de saída
y_{in_k} -	Sinal do k -ésimo neurônio de saída
y_k -	Sinais de entrada
x_i -	

LISTA DE SIGLAS

ANN-	<i>Artificial Neural Network</i>
CAVOK-	<i>Ceiling and Visibility OK</i>
CGNA-	Centro de Gerenciamento de Navegação Aeroportuária
Dir-	Dados meteorológicos de direção do vento
EMA-	Estação Meteorológica de Altitude
EMS-	Estação Meteorológica de Superfície
FAB-	Força Aérea Brasileira
FSI-	<i>Fog Stability Index</i>
GPH500-	Altura geopotencial do nível de pressão de 500 hPa
GPH850-	Altura geopotencial do nível de pressão de 850 hPa
GRNN-	<i>General Regression Neural Network</i>
ICEA-	Instituto de Controle do Espaço Aéreo
IEPV-	Impresso Especial de Proteção ao Voo
INFRAERO-	Empresa Brasileira de Infraestrutura Aeroportuária
ISC-	Índice de Sucesso Crítico
LEPS-	<i>Local Ensemble Prediction System</i>
METAR-	Código meteorológico de informações de superfície
MLP-	<i>Multilayer Perceptron</i>
MM5-	Quinta geração do Modelo de Mesoescala da Universidade do Centro Nacional de Pesquisas Atmosféricas do estado da Pensilvânia
NP-NO-	Fenômeno não previsto e não observado
NP-O-	Fenômeno não previsto e observado
P-NO-	Fenômeno previsto e não observado
P-O-	Fenômeno previsto e observado
PC-	Proporção Correta
PD-	Probabilidade de Detecção
PNN-	<i>Probabilistic Neural Network</i>
PO-	Temperatura do ponto de orvalho
PO500-	Temperatura do ponto de orvalho no nível de pressão de 700 hPa

PO700-	Temperatura do ponto de orvalho no nível de pressão de 700 hPa
Precip-	Dados meteorológicos de precipitação
QFE-	Pressão atmosférica ao nível da estação
QFF-	Pressão atmosférica ajustada ao nível do mar
QNH-	Pressão atmosférica ajustada ao nível padrão do mar
R1-	Rede Neural Artificial para a previsão de uma hora
R2-	Rede Neural Artificial para a previsão de duas horas
R3-	Rede Neural Artificial para a previsão de três horas
RBF-	<i>Radial Basis Function</i>
REDEMET-	Rede de Meteorologia Aeronáutica
RFA-	Razão de Falso Alarme
RNA-	Rede Neural Artificial
TAF-	Código meteorológico de previsão para aeródromo
Temp-	Dados meteorológicos de temperatura do ar
TEMP-	Código meteorológico de informações de altitude
Temp500-	Temperatura do ar no nível de pressão de 500 hPa
Temp700-	Temperatura do ar no nível de pressão de 700 hPa
Vel250-	Velocidade do vento no nível de pressão de 250 hPa
UR-	Dados meteorológicos de umidade relativa
UR500-	Umidade Relativa no nível de pressão de 500 hPa
UR700-	Umidade Relativa no nível de pressão de 700 hPa
UTC-	Horário internacional padrão
V-	Viés
Vel-	Dados meteorológicos de velocidade do vento
Vis-	Dados meteorológicos de visibilidade
WEBMET-	Sistema Automatizado de Registro e Gerenciamento das Observações Meteorológicas

SUMÁRIO

1	INTRODUÇÃO	1
2	REVISÃO BIBLIOGRÁFICA	4
2.1	NEVOEIRO	4
2.1.1	Nevoeiro de Radiação	4
2.1.2	Nevoeiro de Advecção	5
2.1.3	Nevoeiro Orográfico	5
2.2	REDES NEURAIS ARTIFICIAIS	5
2.2.1	O Neurônio Biológico	6
2.2.2	O Neurônio Artificial	7
2.2.3	Redes Neurais Artificiais de Função de Base Radial (RNA RBF)	11
2.3	ESTUDOS SOBRE A APLICAÇÃO DE RNA EM METEOROLOGIA	18
2.4	ESTUDOS SOBRE A PREVISÃO DE VISIBILIDADE UTILIZANDO OUTROS MÉTODOS	20
3	METODOLOGIA	22
3.1	CARACTERÍSTICAS DO AEROPORTO AFONSO PENA	22
3.2	BASE DE DADOS METEOROLÓGICOS	24
3.2.1	Tratamento dos Dados	26
3.2.1.1	Controle de qualidade dos dados	27
3.2.1.2	Pré-processamento dos dados	28
3.2.1.2.1	Pré-processamento das informações de direção do vento	28
3.2.1.2.2	Pré-processamento das informações de visibilidade	29
3.2.1.3	Compatibilização da amostra dos dados de altitude	29
3.2.2	Base de dados tratada	30
3.3	MÉTRICAS PARA AVALIAÇÃO DE DESEMPENHO DE PREVISÕES DE VARIÁVEIS CATEGÓRICAS	31
3.3.1	Medidas escalares de desempenho	32
3.3.1.1	Proporção Correta (PC)	32
3.3.1.2	Índice de Sucesso Crítico (<i>Threat Score</i> - ISC)	33
3.3.1.3	Viés (V)	33

3.3.1.4	Razão de Falso Alarme (RFA).....	34
3.3.1.5	Probabilidade de Detecção (<i>Hit Rate</i> - PD).....	34
3.4	CLIMATOLOGIA E PERSISTÊNCIA.....	36
3.4.1	Climatologia	36
3.4.1.1	Distribuição dos eventos por mês e horário.....	36
3.4.1.2	Distribuição dos eventos por faixas de temperatura do ar.	38
3.4.1.3	Distribuição dos eventos por faixas de umidade relativa...	40
3.4.1.4	Distribuição dos eventos em relação à cobertura de nuvens	41
3.4.1.5	Distribuição dos eventos em relação à pressão atmosférica	43
3.4.1.6	Distribuição dos eventos em relação à direção e à velocidade do vento	44
3.4.1.7	Distribuição dos eventos em relação à precipitação	47
3.4.1.8	Definição das previsões através de climatologia	48
3.4.2	Persistência	49
3.5	ESPECIFICAÇÕES DA REDE NEURAL UTILIZADA	51
3.6	SELEÇÃO DE DADOS DE ENTRADA PARA A RNA	53
4	EXPERIMENTOS DE PREVISÃO DE OCORRÊNCIA DE NEVOEIRO	55
4.1	PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DE CLIMATOLOGIA	55
4.2	PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DE PERSISTÊNCIA	56
4.3	PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DE RNA RBF	57
4.3.1	Treinamento das Redes Neurais Artificiais	57
4.3.1.1	Primeira etapa do treinamento: Treinamento com as séries temporais de visibilidade	58
4.3.1.2	Segunda etapa do treinamento: Primeiro ajuste do número de neurônios ocultos	61
4.3.1.3	Terceira etapa do treinamento: Primeira inserção de novas variáveis	63

4.3.1.4	Quarta etapa do treinamento: Segundo ajuste do número de neurônios ocultos	66
4.3.1.5	Quinta etapa do treinamento: Ajuste da amostra de dados de treinamento	69
4.3.1.6	Sexta etapa do treinamento: Terceiro ajuste do número de neurônios ocultos	73
4.3.2	Operacionalização das Redes Neurais Artificiais ótimas ..	75
4.4	DISCUSSÃO DOS RESULTADOS	77
5	CONCLUSÃO	79
	REFERÊNCIAS	81
	ANEXO A- Correlação linear entre as diversas variáveis e a variável visibilidade	A1
	ANEXO B- Evolução do erro para as Redes R1, R2 e R3 em cada fase de treinamento	A2

1 INTRODUÇÃO

O Aeroporto Afonso Pena, localizado em São José dos Pinhais, Paraná, foi classificado pela Empresa Brasileira de Infraestrutura Aeroportuária (INFRAERO), com base no Anuário Estatístico Operacional de 2015, como o sexto maior do país em movimentação de aeronaves, com cerca de 207 pousos e decolagens diárias. No entanto, em todo o país, segundo o mesmo Anuário, é o aeroporto que mais tem estas atividades interrompidas por restrição de visibilidade, causando grande prejuízo aos passageiros, que muitas vezes têm seus voos atrasados ou cancelados e às empresas aéreas, que acabam por ter despesas extras como com a alocação dos passageiros, reabastecimento das aeronaves, ocupação de pátio, entre outras; além de sobrecarregar o trabalho do Centro de Gerenciamento de Navegação Aérea (CGNA), cujo objetivo é gerenciar o fluxo de tráfego aéreo nacional.

Conseguir prever a formação, o tempo de duração, e a dissipação deste fenômeno contribui sobremaneira na organização da malha aérea do país, auxiliando empresas de transporte aéreo na otimização de tempo e custos, elevando a qualidade dos serviços prestados, além de aumentar a segurança nos pousos e decolagens, pois como bem salientou Varejão-Silva (2001) “nevoeiros sempre caracterizam uma condição de risco em relação ao transporte terrestre e à navegação marítima e aérea”.

Atualmente, a previsão deste fenômeno é realizada de forma subjetiva, por um meteorologista, através de estudos climatológicos, de modelos numéricos como o ETA, ou através de equações probabilísticas como o índice FOG e o *Fog Stability Index* (FSI). Entretanto, apesar da disponibilidade dessas ferramentas, a previsão deste fenômeno está entre os problemas mais difíceis com que se deparam previsores ao redor do mundo (BALLARD, GOLDING e SMITH, 1991).

Por ser um fenômeno meteorológico de formação local e gênese complexa, a previsão de ocorrência e evolução dos nevoeiros é um grande desafio aos meteorologistas e uma grande motivação aos pesquisadores, haja vista as diversas pesquisas, nas mais diversas áreas, que buscam métodos e técnicas de previsão do fenômeno.

As Redes Neurais Artificiais (RNA) são uma poderosa ferramenta computacional que permite o aprendizado de máquina, ou seja, o programa reconhece os padrões exibidos em um banco de dados, e consegue prever o resultado quando novas informações lhe são apresentadas. O emprego desta tecnologia na previsão de nevoeiros pode produzir bons resultados, pois existe um modelo físico na gênese do fenômeno, mas sua complexidade dificulta a análise através de ferramentas tradicionais.

Das diversas configurações de Redes Neurais Artificiais, as RNAs de Funções de Base Radial (*Radial Basis Function*- RBF) se destacam na capacidade de classificação (SILVA, SPATTI e FLAUZINO, 2010). Por esta razão optou-se por esta estrutura de RNA, pois a previsão do fenômeno nevoeiro será tratada nesta pesquisa como classe (ocorrência do fenômeno) e não-classe (não ocorrência do fenômeno).

Esta pesquisa objetiva avaliar o desempenho do uso de Redes Neurais Artificiais de Função de Base Radial na previsão de curtíssimo prazo do fenômeno meteorológico nevoeiro. Tomar-se-á por curtíssimo prazo o período compreendido por até três horas subsequentes, e tratar-se-á da previsão do fenômeno no formato categórico, considerando apenas a ocorrência ou a não ocorrência da condição. Para isso, se busca verificar quais variáveis locais, tais como dados de superfície, influenciam de maneira mais significativa no desempenho da Rede Neural na previsão deste fenômeno; se informações sinóticas, tais como as de ar superior, influenciam no desempenho da RNA, e quais destas informações são mais significativas, destacando que os dados serão usados em sua forma original, e sem pós-processamento. Avaliam-se também quantas horas de dados antecedentes são importantes para o bom desempenho da Rede Neural.

No âmbito da meteorologia aeronáutica, uma das mensagens que pode apresentar a previsão do fenômeno nevoeiro é a mensagem TAF (*Terminal Aerodrome Forecast*). Esta mensagem apresenta a previsão meteorológica para 12, 24 ou 30 horas, de acordo com o porte do aeródromo. No entanto, esta informação deve ser divulgada com, pelo menos, uma hora de antecedência em relação ao início de sua validade, e a previsão do fenômeno pode ocorrer em qualquer período dentro da validade da mensagem. Esta disparidade no horário dos dados utilizados, bem como o período para qual é

feita a previsão impossibilitam a comparação entre o método proposto nesta pesquisa e o atualmente utilizado. Assim, foram utilizadas as técnicas de previsão por climatologia e persistência na verificação do desempenho da utilização de RNAs RBF na previsão do fenômeno nevoeiro.

No Capítulo 2, é realizada uma revisão bibliográfica, com a apresentação dos diversos conceitos que serão utilizados nesta pesquisa, bem como outras pesquisas já realizadas sobre este assunto; no Capítulo 3 é apresentada a metodologia utilizada, com a caracterização do Aeroporto em questão, a apresentação do banco de dados, e a configuração da Rede Neural Artificial utilizada; no Capítulo 4 são apresentados os experimentos realizados e as discussões acerca destes; e as conclusões seguem no Capítulo 5.

2 REVISÃO BIBLIOGRÁFICA

2.1 NEVOEIRO

Segundo Oliveira, Vianello e Ferreira (2001) “Nevoeiro é uma formação de grande quantidade de gotículas de água (excepcionalmente, cristais de gelo), e massa de ar estável, sobre a superfície, provocando redução de visibilidade”. Para fins aeronáuticos, esta restrição deve ser inferior a 1000 metros (MCA 105-2, 2015).

Fisicamente, o nevoeiro se forma quando uma porção de ar alcança a saturação, e o vapor d’água nela contida condensa-se. Existem quatro mecanismos que permitem este processo: o ar é resfriado até alcançar o ponto de orvalho, mas seu volume permanece constante; o volume da porção é aumentado, mas sem o acréscimo de calor (o que leva ao resfriamento do ar, pois a energia é consumida pelo trabalho); alteração conjunta do volume e temperatura, diminuindo a capacidade de retenção de água da porção de ar; ou quando quantidade de vapor d’água na porção de ar (BARRY e CHORLEY, 2003).

Existem diferentes tipos de nevoeiros, classificados de acordo com seu processo de formação. A saber, nevoeiro de radiação, de advecção e orográfico. (VAREJÃO-SILVA, 2001).

2.1.1 Nevoeiro de Radiação

Os nevoeiros de radiação são assim chamados, pois sua gênese é devida ao resfriamento do solo através da liberação do calor da radiação solar e atmosférica acumulados durante o dia. A massa de ar mais próxima ao solo passa a se resfriar, isobaricamente, através da troca de energia com o solo, por condução. No entanto, a quantidade de vapor d’água existente no início do processo continua o mesmo, mas à medida que a porção de ar resfria, esta se aproxima da saturação. Se a porção de ar for suficientemente úmida, ou o resfriamento for suficiente para que a temperatura alcance o ponto de orvalho, ocorrerá a condensação e, por conseguinte, o nevoeiro.

Nevoeiros de radiação se formam em noites sem nuvens, ou com uma camada de nuvens que não gere efeito estufa, de modo a impedir o rápido resfriamento do solo. Portanto, para a formação do nevoeiro de radiação é necessária umidade relativa elevada, grande resfriamento da superfície, e ventos fracos que gerem alguma turbulência de modo que a camada resfriada seja ampliada.

2.1.2 Nevoeiro de Advecção

Este tipo de nevoeiro forma-se a partir de uma massa de ar quente e úmida quando esta escoa por sobre uma superfície mais fria. A massa de ar passa a resfriar-se progressivamente, e quando atinge o ponto de orvalho, forma o nevoeiro. Neste caso, é necessária uma grande diferença de temperatura entre a massa de ar e a superfície, e que a massa esteja se deslocando em baixa velocidade, de modo que possa resfriar-se no processo de troca de calor com a superfície.

2.1.3 Nevoeiro Orográfico

O nevoeiro orográfico ocorre quando uma porção de ar úmido se eleva ao longo de uma encosta. Esta porção passa a resfriar-se adiabaticamente (supondo que o movimento ascendente da porção de ar é suficientemente rápido para que não haja troca de calor com o ambiente). Se o aclave for suficientemente longo e inclinado, a partir da altura em que o ar atingir o ponto de saturação, até o topo da elevação, haverá a formação de nevoeiro.

2.2 REDES NEURAIS ARTIFICIAIS

O termo “Redes Neurais Artificiais” advém do fato de que estas tecnologias buscam representar matematicamente as estruturas cerebrais biológicas. Da mesma maneira que os seres humanos buscam nos eventos passados os auxílios necessários para a resolução de novos problemas, as

redes neurais procuram padrões, em conjunto de dados, que auxiliem em classificações e previsões (ALMEIDA, 2009).

Basicamente, a Rede Neural Artificial detecta os padrões existentes no conjunto de dados utilizado para o treinamento da rede, “aprendendo” como eles se comportam, e quando apresentadas novas informações, ela deve ser capaz de tratá-los de acordo. Segundo Castro e Castro (2002) este método se mostra mais eficaz do que métodos estatísticos convencionais (tais como regressão linear ou análise de variância) quando o processo que rege os dados é não-linear, não-estacionário, ou desconhecido.

2.2.1 O Neurônio Biológico

É sabido que o cérebro biológico é formado por pequenos processadores simples, os neurônios, cuja principal função é conduzir impulsos elétricos. A célula neurológica pode ser dividida em três partes principais: dendritos, corpo celular e axônio, conforme ilustrado na Figura 1, a seguir.

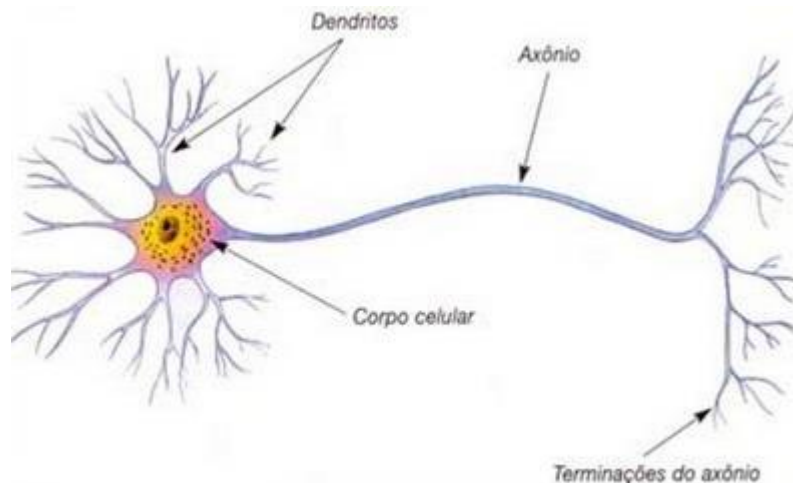


FIGURA 1- REPRESENTAÇÃO DE UM NEURÔNIO BIOLÓGICO
FONTE: Adaptado de Representação de um neurônio biológico (2016).

A função principal dos dendritos é captar, continuamente, os estímulos advindos de outros neurônios, ou, no caso dos neurônios sensitivos, do meio externo. O corpo celular é responsável pelo processamento das informações recebidas pelos dendritos, e possui o potencial de ativar o impulso elétrico através do axônio. O axônio, por sua vez, possui a finalidade de conduzir os

impulsos disparados pelo corpo celular, a outros neurônios ou diretamente aos tecidos, através de suas terminações sinápticas.

As conexões que permitem a transferência de impulsos elétricos entre os neurônios, sem que haja contato entre eles, são chamadas sinapses. Esta transferência de sinal é realizada através de neurotransmissores, os quais ponderam a intensidade do impulso elétrico no processo de transmissão e recepção dos sinais. São, portanto, os neurotransmissores os grandes responsáveis por toda a operação cerebral, pois são eles que ativam ou inibem a ação dos impulsos elétricos entre os neurônios (SILVA, SPATTI e FLAUZINO, 2010).

2.2.2 O Neurônio Artificial

Seguindo a ideia de que as Redes Neurais Artificiais tiveram inspiração nas estruturas cerebrais biológicas, podem-se comparar as unidades processadoras de ambas as estruturas. A unidade computacional de processamento das Redes Neurais Artificiais denomina-se neurônio artificial.

Os neurônios artificiais realizam tarefas muito parecidas com as dos neurônios biológicos, pois coletam sinais de entrada, processam-nos através de uma função aditiva, e produzem uma resposta como saída, através de uma função de ativação. Segundo Silva, Spatti e Flauzino (2010), cada neurônio artificial é constituído dos seguintes elementos básicos:

- a) Sinais de entrada (x_1, x_2, \dots, x_n): são as informações apresentadas à Rede Neural;
- b) Pesos sinápticos (w_1, w_2, \dots, w_n): são os valores que ponderarão as variáveis de entrada, permitindo quantificar a relevância de cada informação à funcionalidade do neurônio. No processo de treinamento, a RNA ajusta os pesos de cada sinal de entrada, de forma que alcance o melhor desempenho possível. A este processo chamamos Aprendizagem; Realizam o trabalho dos neurotransmissores, na comparação ao modelo biológico;
- c) Combinador linear (Σ): efetua o somatório dos sinais de entrada já ponderados;

- d) Limiar de ativação (θ): é a variável que determina o parâmetro para que o resultado produzido pelo combinador linear possa gerar um valor de saída do neurônio;
- e) Potencial de ativação (u): é o resultado obtido através da subtração entre o valor gerado pelo combinador linear, e o limiar de ativação. Valores de u positivos têm potencial excitatório, ao passo que, valores negativos têm potencial inibitório;
- f) Função de ativação (g): produz o resultado de saída do neurônio, limitando-o, ou não, a um intervalo de valores que constituem seu conjunto imagem. As principais funções utilizadas como função de ativação são a função degrau, a função degrau bipolar, a função rampa simétrica, a função logística, a função tangente hiperbólica, a função gaussiana e a função identidade.
- g) Sinal de saída (y): é o sinal final produzido pelo neurônio artificial dado por

$$y = g(u) = f\left(\sum_{i=1}^n w_i x_i + \theta\right) \quad (1)$$

A Figura 2, abaixo, representa um neurônio artificial, com os elementos descritos anteriormente.

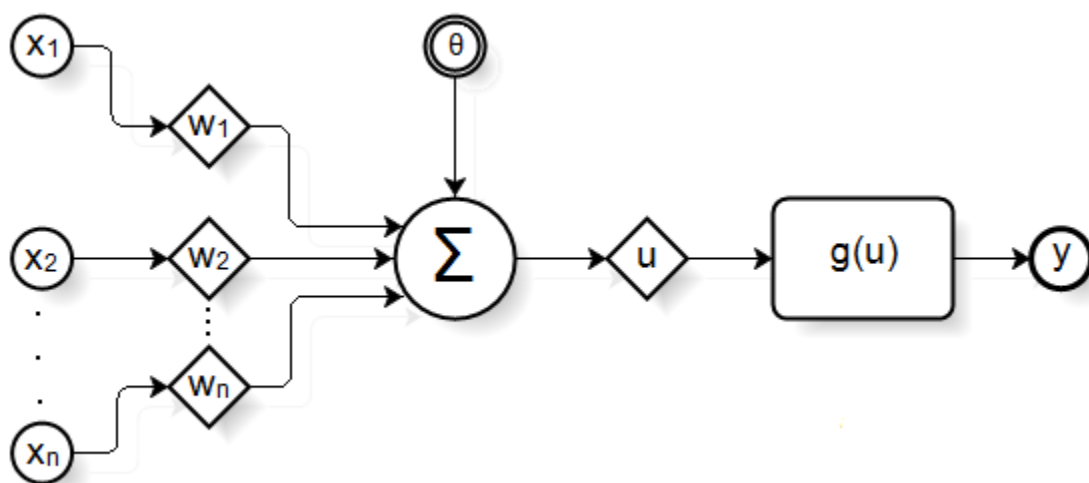


FIGURA 2- REPRESENTAÇÃO UM NEURÔNIO ARTIFICIAL
 FONTE: A Autora, 2016.

Embora cada neurônio seja simples, a grande capacidade computacional das RNAs está na combinação de diversos destes pequenos processadores, formando as redes de conexões que dão nome a essa tecnologia. Esta associação de neurônios é que possibilita às RNAs a resolução de problemas bastante complexos, com a análise de muitas variáveis em conjunto. Mas, para alcançar esta capacidade, a RNA deve passar pelo processo chamado aprendizagem, o que lhe permitirá detectar os padrões existentes do conjunto de dados, possibilitando a operação da Rede Neural.

O processo de aprendizagem pode ser classificado como supervisionado ou não-supervisionado. Na aprendizagem supervisionada são apresentados à RNA os valores de entrada e os de saída (referências), o que orienta a rede em relação ao que se espera dela. Pode-se, ainda, subdividir este grupo, de acordo com a regra que realiza o ajuste dos pesos das conexões. Algumas destas regras são:

- a) Aprendizagem por correção do erro: se busca ajustar os pesos, de modo que o erro obtido da diferença entre o valor de saída gerado pela rede, e o valor de saída oferecido como referência seja o mínimo possível.
- b) Aprendizagem por reforço: sinais de reforço ou penalidade são emitidos a cada sinal produzido pela RNA.

Na aprendizagem não-supervisionada, os dados não são apresentados na saída, apenas na entrada. A RNA decide por si só quais as relações mais fortes existentes entre as variáveis de entrada. Pode-se também subdividir este grande grupo, de acordo com a regra que coordena a aprendizagem. Eis alguns exemplos:

- a) Aprendizagem competitiva: os neurônios competem entre si para determinar qual deles responde melhor a certo estímulo;
- b) Aprendizagem Hebbiana: baseia-se no postulado de Hebb, o qual define que se a ativação de um neurônio influencia na ativação de outro, esta ligação deve ser reforçada. Caso contrário esta ligação deve ser enfraquecida;
- c) Aprendizagem baseada na memória: neste caso, padrões são armazenados, e recuperados de acordo com os estímulos liberados.

Em relação à organização das camadas e ao número de neurônios, existem diversas topologias para as Redes Neurais Artificiais. Algumas podem ter camadas de neurônios entre a camada de entrada e a de saída, as quais se denominam camadas ocultas. Em relação à propagação do sinal, algumas RNAs enviam o sinal dos neurônios sempre adiante, para a próxima camada de neurônios. Essas configurações são conhecidas como *feedforward*. Caso haja propagação de sinal também no sentido inverso, a rede é chamada de recorrente (*backward*).

Algumas das estruturas mais comuns são as seguintes:

- a) Perceptron Multicamadas (*Multilayer Perceptron*- MLP): Possui uma ou mais camadas ocultas e o sinal sempre é propagado adiante.
- b) Rede de Base Radial (*Radial Basis Function*- RBF): Possui uma camada oculta cujos nós possuem funções de base radial.
- c) Redes Construtivas- A estrutura da rede inicia pequena, mas ganha volume conforme o treinamento evolui.
- d) Autorganizáveis (*Self-Organizing*): Baseada na aprendizagem competitiva.
- e) Redes de Jordan e Elman: Redes parcialmente recorrentes. Na Rede de Elman a realimentação dá-se entre os neurônios da mesma camada. Na de Jordan, a recorrência acontece da camada de saída para a de entrada.
- f) Redes de Hopfield: São redes autoassociativas recorrentes. Armazenam padrões na memória, que são recuperados conforme os estímulos dos sinais de entrada.
- g) Redes Híbridas: Combinam RNA e outro subsistema qualquer, como Métodos Estatísticos ou Algoritmos Genéticos.

Quanto à aplicabilidade desta tecnologia, existem diversas áreas nas quais o uso de RNAs pode proporcionar bons resultados, tais como aproximação universal de funções, controle de processos, reconhecimento e classificação de padrões, agrupamento de dados, sistemas de previsão, otimização de sistemas, e memórias associativas, entre outras (SILVA, SPATTI e FLAUZINO, 2010).

As RNAs vêm sendo particularmente utilizadas no campo da Meteorologia, dada sua capacidade de identificação de padrões complexos no

trabalho com muitas variáveis. A análise e a previsão de séries temporais é um assunto bastante difícil de ser tratado por métodos tradicionais, e as RNAs vêm como uma poderosa ferramenta no auxílio a estas análises mais complexas.

2.2.3 Redes Neurais Artificiais de Funções de Base Radial (RNA RBF)

Funções de Base Radial são funções com domínio real, cuja saída depende da distância entre o ponto de entrada x , e um ponto central c , determinado. Ou seja,

$$\phi(x) = \phi(||x - c||) \quad (2)$$

onde $\phi(x)$ é a RBF.

Assim, pode-se usar, como função de ativação, qualquer função de distância que satisfaça o disposto acima. Estas funções geralmente são do tipo gaussianas, a saber,

$$g(x) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (3)$$

onde x é o vetor apresentado, c é o centro da função gaussiana, e σ é o desvio padrão da função. Assim, $g(x)$ depende de quão distante está o valor de x em relação ao centro c da gaussiana. Os parâmetros ajustáveis em uma função gaussiana são a posição do centro c e o desvio padrão σ . Quanto maior o desvio padrão, mais alargada será a base da gaussiana.

A figura a seguir representa o gráfico de uma função gaussiana.

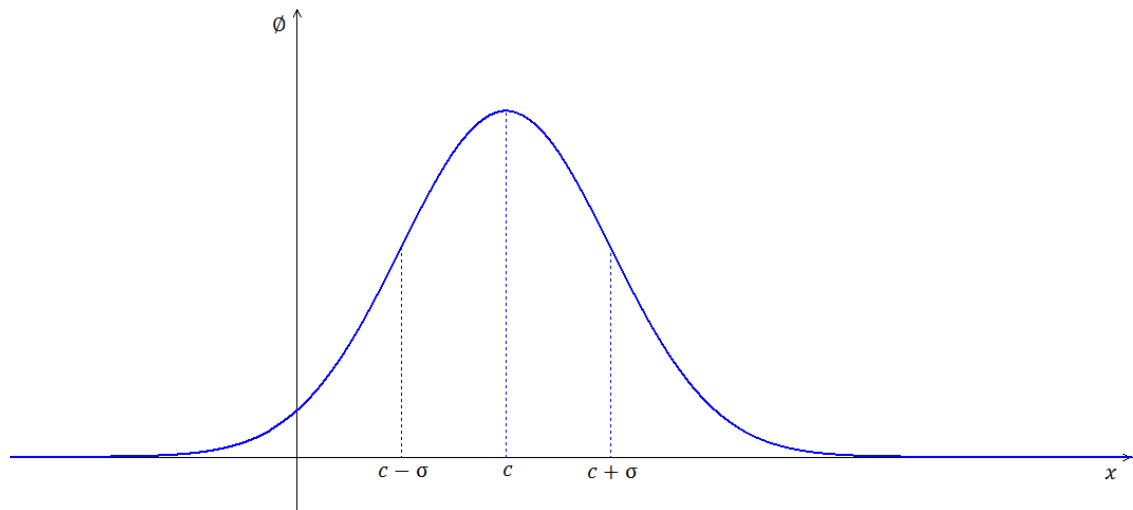


FIGURA 3- FUNÇÃO GAUSSIANA
FONTE: A Autora, 2016.

RNAs RBF estão dentre as RNA do tipo *feedforward*, e possuem três camadas: uma para receber os vetores com dados de entrada; uma camada oculta, onde cada neurônio possui uma RBF como função de ativação; e uma camada para os valores de saída, que pode conter um ou mais neurônios. Um esquema representando a operação de uma RNA RBF pode ser observado na Figura 4.

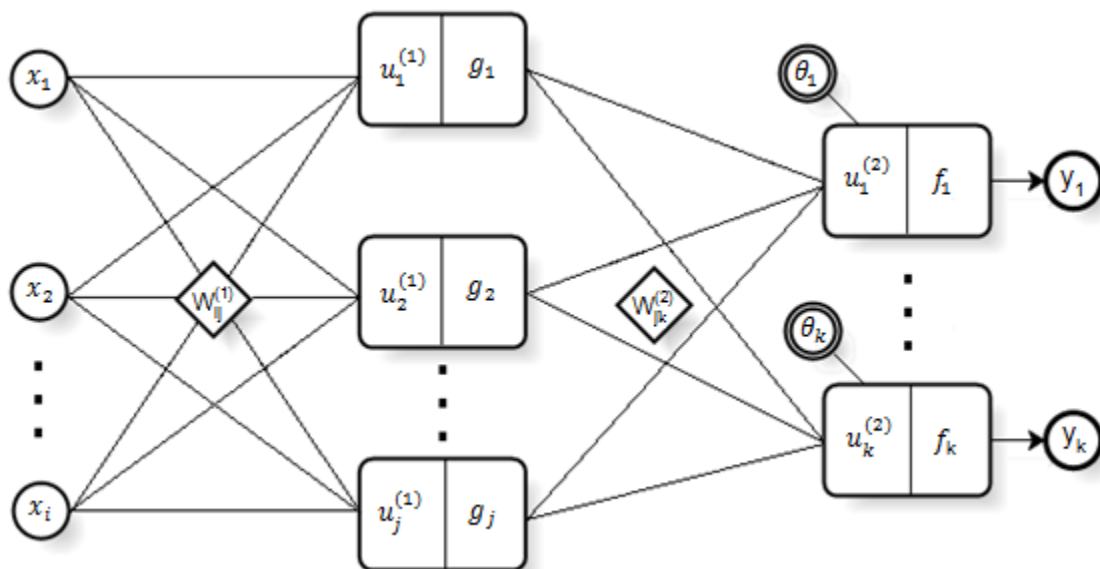


FIGURA 4- REPRESENTAÇÃO DA OPERAÇÃO DE UMA RNA RBF
FONTE: A Autora, 2016.

Na Figura 4 x_i são os vetores de entrada; $W_{ij}^{(1)}$ são os pesos sinápticos da ligação entre o neurônio i e o neurônio oculto j ; g_j são as funções de ativação da camada oculta (a saber, funções gaussianas); $W_{jk}^{(2)}$ são os pesos sinápticos que ponderam as ligação entre o neurônio oculto j e o neurônio de saída k e f_k são as funções de ativação da camada de saída.

Além da função de ativação de base radial, outra característica importante das redes RBF é o processo de treinamento no ajuste dos centros e desvios padrões das gaussianas dos neurônios ocultos e dos pesos sinápticos entre os neurônios ocultos e os de saída.

Este processo consiste em duas etapas. A primeira etapa é do tipo não-supervisionada. Nela são ajustados os pesos das conexões entre os neurônios de entrada e da camada intermediária, alocando as funções de base radial de cada um dos neurônios ocultos.

Inicialmente, define-se o número de neurônios da camada oculta. Um vetor representante central é escolhido, aleatoriamente, para cada um dos neurônios ocultos. Ele indicará o centro da curva gaussiana. Cada neurônio oculto formará uma classe em torno dos valores de cada um desses centros iniciais.

Para o caso de classificação de padrões, a RNA RBF considera a distância radial de cada elemento em relação ao centro da gaussiana, construindo hiperesferas delimitadoras em cada neurônio oculto, as quais definirão classes. Assim, um elemento quando apresentado a uma classe a qual não pertence, estará distante do centro da hiperesfera, a ponto de produzir um sinal muito pequeno; ao passo de que quando apresentado à sua classe representante, ficará mais próximo ao centro, produzido um sinal maior.

Desta forma, um bom ajuste dos centros das gaussianas dos neurônios intermediários é fundamental para o bom desempenho da RNA. Uma das melhores maneiras de realizar este ajuste, segundo Silva, Spatti e Flauzino (2010), é através do método de k -médias. Este método consiste no ajuste do centro através da média da distância entre a posição atual do centro da classe e os elementos que a compõem. Os valores dos centros são iniciados aleatoriamente. Quando um vetor é apresentado, calculam-se as distâncias entre este vetor e cada um dos centros. O que apresentar menor distância é o

vencedor, e o vetor passa a pertencer àquela classe. Após a classificação de todos os vetores, a posição do centro é ajustada tomando-se o ponto médio de cada uma das coordenadas entre a posição anterior do centro e o vetor apresentado. Ajustado o valor do centro, toma-se outro vetor pertencente à classe, e repete-se o processo até que todos os vetores sejam apresentados. Todos os vetores do grupo de treinamento são então novamente apresentados à RNA, e redistribuídos nas classes, agora com novos centros. Repete-se este processo de ajuste dos centros até que estes não se movimentem mais do que um valor pré-estabelecido. Este valor deve ser pequeno a ponto de significar a quase estagnação do processo de ajuste.

Quando os vetores centrais já estão bem definidos, a RNA calcula a variância dos valores que ficaram em cada classe, atribuindo-lhes uma curva gaussiana. Cada classe terá uma gaussiana correspondente, que terá a base tão estreita, quanto mais aglomerados estiverem os vetores de treinamento em torno do vetor central. O sinal produzido então, por cada neurônio oculto, quando lhe for apresentado um vetor de entrada será dado por

$$\varphi_j = g_j(x_i) = e^{-\frac{(x_i - c_j)^2}{2\sigma_j^2}} \quad (4)$$

onde c_j refere-se ao representante central do j -ésimo neurônio oculto.

A Tabela 1 e a Figura 5, abaixo, apresentam um esquema dos procedimentos computacionais adotados nesta primeira fase do treinamento da Rede RBF em forma de pseudocódigo e fluxograma, respectivamente.

TABELA 1- ESQUEMA DOS PROCEDIMENTOS COMPUTACIONAIS PARA A PRIMEIRA FASE DO TREINAMENTO

<p><1> Determinar os dados para amostra de treinamento</p> <p><2> Iniciar o valor dos centros dos neurônios da camada intermediária com valores aleatórios</p> <p><3> Repetir quantas vezes for necessário</p> <p> <3.1> Operar com todos os vetores da amostra</p> <p> <3.1.1> Calcular a distância entre o vetor e cada um dos representantes centrais de cada classe</p> <p> <3.1.2> Selecionar o neurônio oculto que apresenta menor distância ao vetor</p> <p> <3.1.3> Atribuir o vetor a esta classe</p>
--

<3.2> Operar com os representantes centrais de todas as classes

<3.2.1> Ajustar a posição dos representantes centrais

<3.2.2> Se a movimentação das posições dos centros (medida pela diferença na posição antes e após a apresentação dos vetores) for suficientemente pequena, ir para <4>, se não, voltar para <3>

<4> Calcular as variâncias de cada classe, atribuindo-lhe uma gaussiana

FONTE: A Autora, 2016.

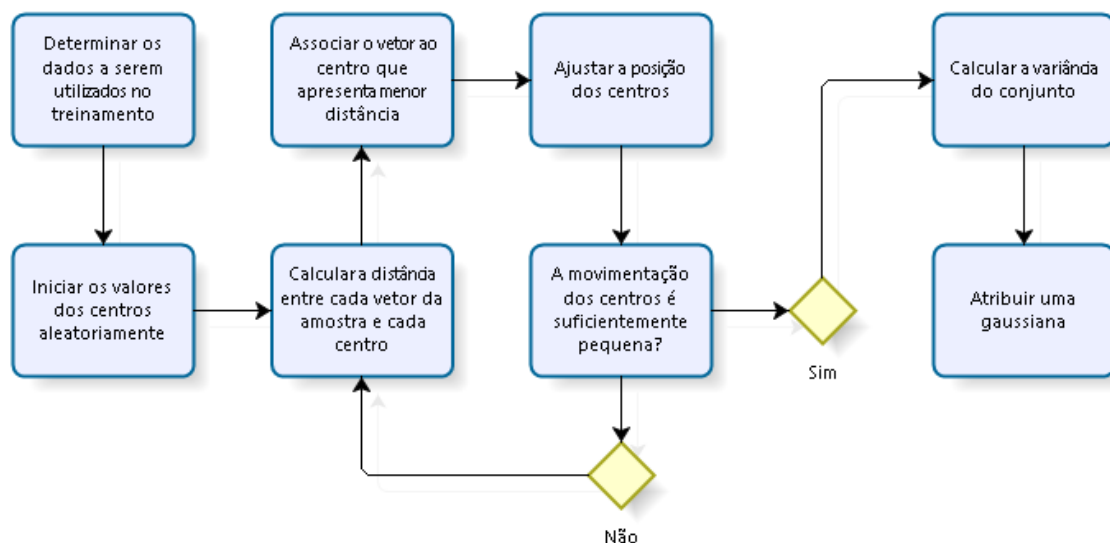


FIGURA 5- FLUXOGRAMA DA PRIMEIRA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

Ao término da primeira fase do treinamento, inicia-se a segunda fase. Nela são ajustados os pesos das conexões da camada de saída para obter a combinação linear final.

Ao sinal ϕ_j , obtido da aplicação da função de ativação da camada oculta, é aplicado o peso $W_{jk}^{(2)}$ da sinapse que conecta o neurônio j da camada oculta ao neurônio k da camada de saída, produzindo o sinal $w_{jk}\phi_j$. Em cada neurônio de saída é obtido o somatório dos sinais de entrada, produzindo no neurônio de saída k o valor y_{in_k} . A essa soma ponderada é então aplicada função de ativação do neurônio de saída f , produzindo o sinal de saída $y_k = f(y_{in_k})$. Os diversos sinais dos neurônios da camada de saída formam o vetor de saída y .

O vetor y é comparado com o vetor de saída de referência ϵ , formado pelos elementos d_k .

O erro do vetor de saída produzido pela rede neural é definido por

$$E = \frac{1}{2} \sum_k (\epsilon_k - y_k)^2. \quad (5)$$

Conhecido o erro, é então aplicado o método de correção estocástica das posições t_j , das variâncias σ_j^2 , e dos pesos w_{jk} da rede, por meio das seguintes expressões, respectivamente:

$$\begin{aligned} t_j(n+1) &= t_j(n) + \mu_t \left(\frac{u - t_j}{\sigma_j^2} \right) \varphi_j \left(\sum_k \delta_k w_{jk} \right); \\ \sigma_j^2(n+1) &= \sigma_j^2(n) + \mu_\sigma \left(\sum_k \delta_k w_{jk} \right) \frac{\|u - c_j\|^2}{(\sigma_j^2)^2} \varphi_j; \\ w_{jk}(n+1) &= w_{jk}(n) + \mu_w \delta_k \varphi_j \end{aligned} \quad (6)$$

onde, n e $n+1$ são duas iterações sucessivas de correção da rede; μ_c , μ_σ e μ_w são os passos de treinamento das posições, variâncias das funções gaussianas e dos pesos da segunda camada da rede, respectivamente; e $\delta_k = (\epsilon_k - y_k) f'(y_{in})_k$, sendo $f'(y_{in})_k$ a derivada da função de ativação da camada de saída em relação a y_{in}_k .

Após os ajustes, o conjunto de teste é submetido à rede. Este processo gera vetores de saída, os quais são comparados aos vetores de referência, resultando em um valor de erro de teste. Caso este valor seja inferior ao da iteração anterior, é iniciado um novo ajuste (pois esta condição sugere um comportamento decrescente dos valores de erro) nos valores dos pesos e posições dos representantes centrais das classes. Caso o valor do erro de teste aumente, a rede salva a configuração anterior, que gerara erro inferior, e interrompe o processo de ajuste.

A Tabela 2, e a Figura 6, abaixo, apresentam um esquema dos procedimentos computacionais adotados na segunda fase do treinamento da Rede em pseudocódigo e fluxograma, respectivamente.

TABELA 2- ESQUEMA DOS PROCEDIMENTOS COMPUTACIONAIS PARA A SEGUNDA FASE DO TREINAMENTO

<1> Determinar o conjunto de amostras de treinamento
 <2> Determinar o vetor de saída desejado para cada amostra
 <3> Iniciar os pesos sinápticos aleatoriamente
 <4> Determinar a taxa de aprendizagem e o erro máximo
 <5> Repetir para todas as amostras
 <5.1> Obter os valores de saída de cada neurônio intermediário
 <5.2> Propagar o sinal para a camada de saída
 <5.3> Obter o sinal de saída
 <5.4> Comparar o sinal de saída com a referência e obter o erro de treinamento
 <5.5> Ajustar os valores dos pesos das conexões e dos centros e variâncias das RBF's
 <6> Submeter os valores do conjunto de teste à rede. Se o erro do conjunto de teste for maior que o anteriormente obtido, <fim>; se não, ir para <5>

FONTE: A Autora, 2016.

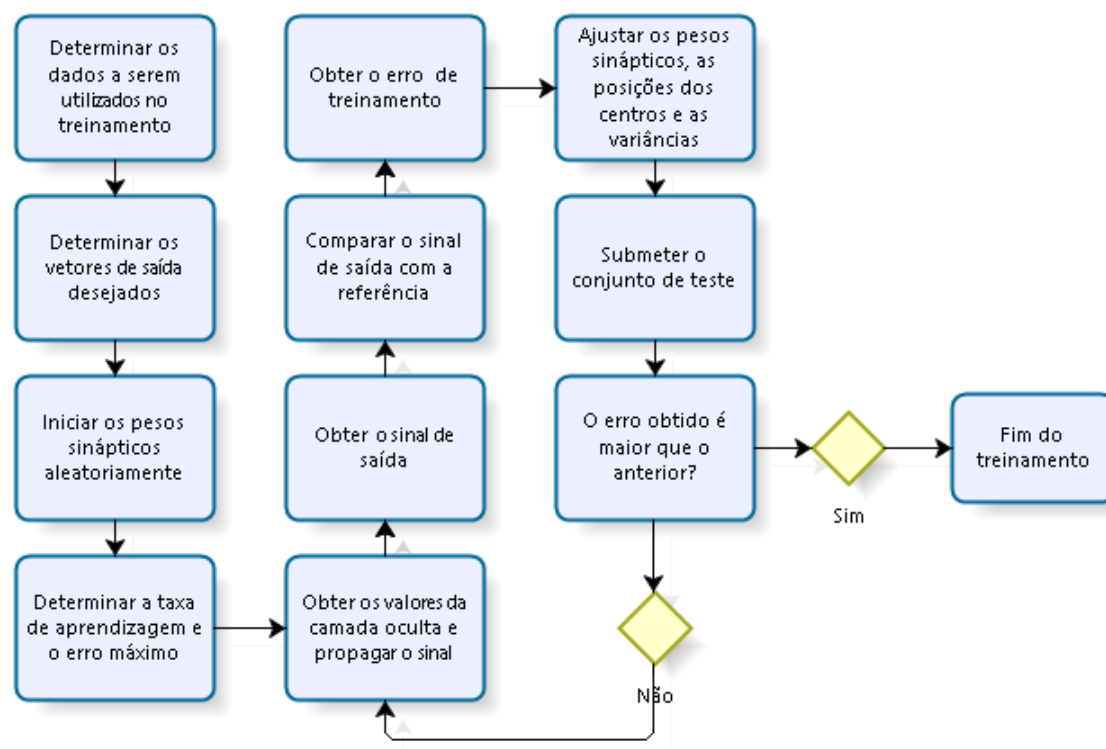


FIGURA 6- FLUXOGRAMA DA SEGUNDA FASE DE TREINAMENTO
 FONTE: A Autora, 2016.

Como a rede inicialmente separa grupos de dados para depois fazer o ajuste local, isso permite que as RNAs RBF aprendam mais rapidamente padrões complexos e se adaptem melhor a mudanças. Esta característica as faz, segundo Castro e Castro (2002), especialmente adequadas para a previsão de séries temporais, particularmente, séries de dados regidas por processos não-lineares e/ou não-estacionários.

2.3 ESTUDOS SOBRE A APLICAÇÃO DE RNA EM METEOROLOGIA

Como já citado, as RNAs têm ganhado destaque nas pesquisas em meteorologia. Diversos autores já utilizaram as mais diversas estruturas e configurações de Redes Neurais para abordar os mais variados assuntos na área. Sobre a utilização de RNA para a previsão de visibilidade, pode-se destacar algumas pesquisas, tais como:

Almeida (2009) comparou os resultados obtidos de duas estruturas de Redes Neurais Artificiais (*Probabilistic Neural Network- PNN* e *General Regression Neural Network- GRNN*) com os da Regressão Linear Múltipla na previsão de até três horas para a previsão de teto e visibilidade para o Aeroporto de Guarulhos, São Paulo. Os resultados da PNN foram os mais significativos, com 92%, 79% e 70% de Proporção Correta para a primeira, segunda e terceira horas de previsão, respectivamente.

Fabbian, Dear e Lellyett (2007) utilizaram uma Rede *Multilayer Perceptron* para prever nevoeiro no Aeroporto de Camberra, Austrália. Os prazos de previsão foram divididos em grupos de 3, 6, 12 e 18 horas. Os autores puderam concluir que a Rede Neural Artificial é adequada para a previsão deste fenômeno no Aeroporto em questão.

Silva (2007) comparou o desempenho de um modelo probabilístico com o de duas Redes Neurais no pós-processamento de um modelo de previsão de visibilidade restrita e ventos fortes, e concluiu que o uso na Rede aumenta a previsibilidade dos eventos em questão.

Kozar, Misra e Powell (2016) utilizaram diversas RNAs no auxílio à previsão da intensidade de ciclones tropicais do Atlântico Norte. Os autores desenvolveram um esquema estatístico-dinâmico chamado Previsão Estatística

de Energia Cinética Integrada, o qual combina um modelo estatístico de previsão de energia cinética a diversas Redes Neurais, permitindo a quantificação do campo de vento em torno do ciclone, possibilitando a análise de sua intensidade e extensão. Segundo eles, a utilização das RNAs no lugar de modelos de regressão linear mais básicos permite ao esquema a distinção de sinais não-lineares, o que resulta em um desempenho significativamente superior em comparação às técnicas de previsão de persistência e climatologia.

Shank, Hoogenboom e McClendon (2008) utilizaram uma Rede Neural Artificial de retropropagação para o desenvolvimento de uma ferramenta de previsão da temperatura do ponto de orvalho para o estado da Geórgia, Estados Unidos. Os autores utilizaram um banco de dados meteorológicos de três anos, coletados em 20 localidades do estado supracitado, e intervalos de previsão de 1, 4, 8 e 12h. A combinação entre variáveis que apresentou melhores resultados incluiu além da própria temperatura do ponto de orvalho, os dados de umidade relativa, radiação solar, temperatura do ar, velocidade do vento e pressão de vapor. Os erros médios absolutos para os intervalos de previsão foram de 0,550°C, 1,234°C, 1,799°C e 2,280°C, respectivamente, e, segundo os autores, predizem adequadamente a variável em questão.

Ainda voltado à Meteorologia, podemos citar outros diversos trabalhos com a utilização de Redes Neurais Artificiais. Rosette (2011) utilizou uma RNA para a previsão de precipitações extremas no Aeroporto do Galeão, Rio de Janeiro; Mathieu e Aires (2016) pesquisaram a utilização de Redes Neurais na extração das informações meteorológicas mais pertinentes e quais seus impactos em modelos estatísticos de previsão; Patil, Deo e Ravichandran (2016) combinaram o resultado de modelos numéricos a RNAs na previsão da temperatura superficial do Oceano Índico; Saha e Mitra (2016) utilizaram Redes Neurais recorrentes na previsão das monções indianas. Segundo eles, a estrutura recorrente da RNA supera tanto o desempenho da estrutura multicamadas, quanto o desempenho de outros modelos numéricos utilizados no Departamento de Meteorologia da Índia, com erro de, aproximadamente, 3,3%.

A aplicabilidade de Redes Neurais Artificiais no campo da Meteorologia pode ser observada nos trabalhos citados. É possível notar a grande

diversidade de temas abordados com o uso desta tecnologia, o que indica o avanço da utilização desta ferramenta computacional nesta área de pesquisa.

2.4 ESTUDOS SOBRE A PREVISÃO DE VISIBILIDADE UTILIZANDO OUTROS MÉTODOS

A previsão das condições de visibilidade é um tema bastante abordado em pesquisas na área de meteorologia, não só através da utilização de Redes Neurais Artificiais, como citado anteriormente, mas também através de diversas outras ferramentas tecnológicas e técnicas de previsão.

Hansen (2007) utilizou Lógica-Fuzzy na previsão de altura de nuvens e visibilidade horizontal para 190 aeroportos canadenses, durante um ano. Este método permitiu medir a similaridade entre séries de dados passados e presentes na projeção de condições para até 24 horas. Utilizando o índice de Heidke, o autor alcançou um desempenho de 56% em previsões para até 6 horas, e 40% para previsões entre 7 e 24 horas.

Roquelaure e Bergot (2009) avaliaram a habilidade do modelo probabilístico *Local Ensemble Prediction System* (LEPS) na previsão de visibilidade e/ou teto restritos no Aeroporto Paris-Charles de Gaulle, Paris, França, durante os invernos dos anos de 2002 a 2007. Os autores observaram que em 78% dos casos, a restrição de teto é causada por nuvens baixas do tipo *stratus*, e as restrições de visibilidade causadas por nevoeiros de advecção e radiação. Para estes casos os autores concluíram que o sistema LEPS é adequado na previsão para curto prazo.

Boneh *et al* (2015) desenvolveram uma ferramenta de apoio à previsão de nevoeiros para o Aeroporto de Melbourne, Victoria, Austrália. Os pesquisadores utilizaram 34 anos de dados meteorológicos no desenvolvimento de um programa objetivo para a previsão de nevoeiros baseado na probabilística Bayesiana. Segundo os autores, após a avaliação da operacionalidade do programa, este se mostrou superior às técnicas anteriormente utilizadas, e passou a ser aceito pelos meteorologistas locais.

Marzban, Leyton e Colman (2007) desenvolveram Redes Neurais Artificiais para pós-processamento de um modelo numérico como auxílio à previsão de visibilidade e teto para 39 aeródromos do noroeste dos Estados

Unidos. A ferramenta combina dados meteorológicos coletados nas estações meteorológicas de superfície aos resultados do modelo numérico MM5 (quinta geração do Modelo de Mesoescala da Universidade do Centro Nacional de Pesquisas Atmosféricas do estado da Pensilvânia). Comparando o desempenho das RNAs com as técnicas de regressão logística e estatísticas de saída do modelo, foi possível observar que o desempenho das Redes Neurais geralmente é superior, dependendo da métrica de avaliação utilizada.

É possível perceber que a previsão das condições meteorológicas de visibilidade despertam o interesse dos pesquisadores, que utilizam as mais diversas ferramentas tecnológicas e técnicas de previsão na busca de índices cada vez melhores. Nota-se que muitos autores, inclusive, utilizam mais de uma ferramenta ou técnica em seus estudos, o que atesta a grande complexidade da previsão desta variável meteorológica.

Em pesquisas realizadas nos sítios da *American Meteorological Society* e Google Acadêmico não foram encontrados trabalhos utilizando RNAs RBF em assuntos relacionados à meteorologia, tampouco na previsão de visibilidade. Este fato mostra que a abordagem apresentada nesta pesquisa virá a acrescentar no conhecimento científico atual. No entanto, a ausência de outras pesquisas com o mesmo enfoque traz dificuldades na avaliação do desempenho atingido, visto que não é possível a comparação com outros estudos similares.

Além disso, outro fator que prejudica a avaliação de desempenho do uso de RNAs RBF em comparação a outras configurações de RNAs é o fato de que, de maneira geral, é perceptível, nos diversos trabalhos pesquisados, certa falta de clareza em relação à identificação das métricas empregadas na avaliação dos desempenhos. Os índices são apresentados, mas em muitos textos não é claro o modo como foi realizada a avaliação. Como será abordado na Seção 2.5, no tocante a fenômenos de baixa frequência de ocorrência (como é o caso dos nevoeiros), uma escolha acertada da métrica de avaliação de desempenho é essencial para a interpretação dos resultados da pesquisa.

3 METODOLOGIA

3.1 CARACTERÍSTICAS DO AEROPORTO AFONSO PENA

O Aeroporto Internacional Afonso Pena está localizado na cidade de São José dos Pinhais, região metropolitana na cidade de Curitiba, no estado do Paraná. Encontra-se em uma região serrana e, segundo o Instituto de Cartografia da Aeronáutica, a 911m de altitude em relação ao nível do mar.

Segundo a INFRAERO (2016), foi originalmente construído com o propósito de servir como base militar para o exército aliado na 2ª Guerra Mundial. Por este motivo, a área escolhida para sua construção tinha alta ocorrência de nevoeiro, pois isso facilitaria na camuflagem das aeronaves.

Em 1946, o aeroporto passou a ser usado pela aviação comercial, recebendo a demanda do Aeroporto do Bacacheri. A partir deste momento, a alta frequência de ocorrência de nevoeiros passou a ser um empecilho às operações aeronáuticas, causando atrasos e cancelamento de voos.

A Figura 7, a seguir, indica a posição da cidade de São José dos Pinhais e o relevo de montanhas próximos à cidade. É possível, ainda, perceber a proximidade com o oceano. Este fator tem relação com a alta frequência de nevoeiros, pois ventos quentes e úmidos vindos desta região configuram condições que favorecerem a formação do fenômeno.

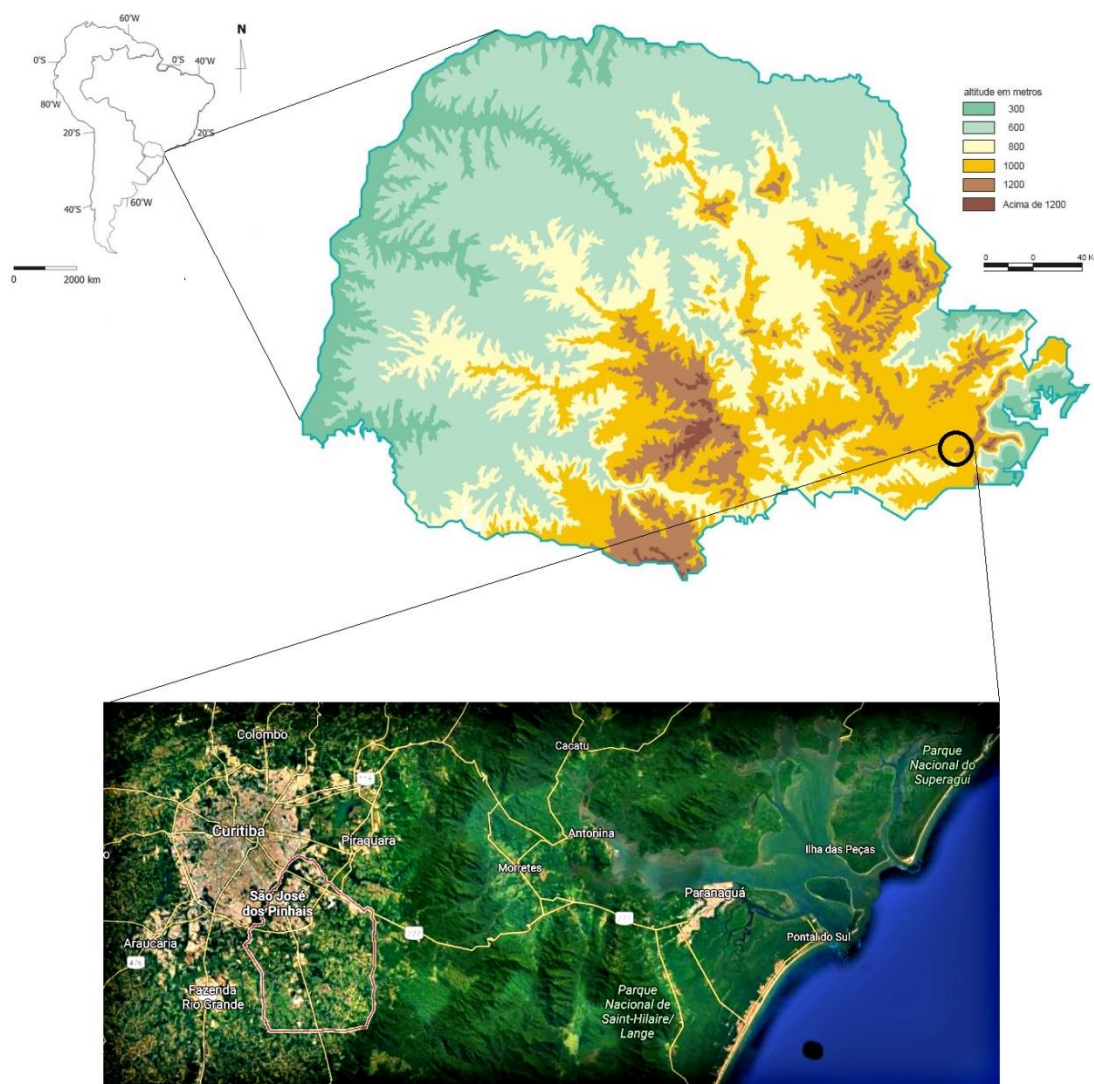


FIGURA 7- OROGRAFIA DAS VIZINHANÇAS DA CIDADE DE SÃO JOSÉ DOS PINHAIS
 FONTE: Adaptado de Mapa do Brasil (2016), Mapa de altitude do estado do Paraná (2016) e Mapa da cidade de São José dos Pinhais (2016).

Por estes motivos, o Aeroporto Afonso Pena permanece impedido de realizar pousos ou decolagens, devido à restrição de visibilidade, aproximadamente, 400 horas por ano. Este quantitativo é mais de três vezes maior do que o apresentado no Aeroporto Internacional Salgado Filho, localizado em Porto Alegre, Rio Grande do Sul, que, segundo dados divulgados por Almeida (2008), apresenta restrição das operações supracitadas por, aproximadamente, 120 horas por ano.

3.2 BASE DE DADOS METEOROLÓGICOS

Os dados meteorológicos utilizados neste estudo descrevem condições de superfície e altitude compreendidas no período de 01/01/1997 00h 00min, a 31/12/2014 23h 59min. Os dados de superfície foram coletados na Estação Meteorológica de Superfície (EMS) de Curitiba, localizada no Aeroporto Internacional Afonso Pena (lat.: 25°31'54"S ; lon.: 049°10'34"W), e obtidos do Sistema Automatizado de Registro e Gerenciamento das Observações Meteorológicas (WEBMET) e do formulário IEPV 105-78, através do Instituto de Controle do Espaço Aéreo (ICEA). As informações de altitude foram coletadas das radiossondagens realizadas na Estação Meteorológica de Altitude (EMA) de Curitiba, localizada no mesmo Aeroporto, através do sítio da Universidade de Wyoming (<http://weather.uwyo.edu/upperair/sounding.html>).

No âmbito da Meteorologia Aeronáutica, as observações meteorológicas à superfície podem ser classificadas como Regulares, Especiais e Locais. As observações regulares são realizadas nas horas cheias, as observações locais quando ocorre algum incidente ou acidente aeronáutico no aeródromo que comporta a Estação Meteorológica de Superfície, e as especiais sempre que houver mudanças significativas nas condições meteorológicas. (MCA 105-2, 2015)

Segundo MCA 105-16 (2015), o Sistema WEBMET foi desenvolvido com o intuito de registrar dados meteorológicos para a formação de um banco de dados que propicie o desenvolvimento de estudos que dinamizem ainda mais os serviços prestados pelo Serviço de Meteorologia Aeronáutica da Força Aérea Brasileira. Sua configuração permite enviar dados de observações meteorológicas de maneira automatizada, mantendo assim, a integridade e confiabilidade dos dados meteorológicos obtidos pelas EMS e EMA. São registrados os seguintes dados meteorológicos nas observações:

- a) direção e velocidade do vento, relacionadas a cabeceira de cada pista de pousos e decolagens, em dezenas de graus e nós, respectivamente;
- b) visibilidade predominante e visibilidade mínima observada na estação, em decâmetros;

- c) alcance visual na cabeceira de cada pista, em metros;
- d) condições de tempo através de código pré-estabelecido;
- e) quantidade (em oitavos), tipo, altura (em decâmetros) e direção das camadas de nuvens;
- f) temperatura do ar, temperatura do ar na pista, temperatura do ponto de orvalho, temperatura do bulbo seco, e temperatura do bulbo úmido com uma casa decimal de precisão;
- g) umidade relativa em unidades inteiras;
- h) pressão atmosférica ao nível da pista (QFE), ajustada ao nível padrão do mar (QNH) e ajustada ao nível do mar (QFF) em hectopascals, com uma casa decimal de precisão;
- i) duração e quantidade de precipitação, em horas e minutos, e em milímetros, respectivamente;
- j) condições de cortante de vento; e
- k) em estações implantadas em plataformas marítimas, temperatura e estado do mar.

A cobertura de nuvens é registrada com numerais de 0 a 8, considerando-se a abóbada celeste dividida em oitavos. Contam-se quantos oitavos do céu estão cobertos por cada tipo de nuvens. Para a informação de cobertura total de nuvens é observado quantos oitavos da abóbada estão cobertos por nuvens, independentemente de suas alturas ou tipos.

Quando na indisponibilidade do Sistema WEBMET, as informações supracitadas são inseridas, de forma manuscrita, no formulário IEPV 105-78, e posteriormente registradas no Sistema. Este formulário contém os mesmos campos de que dispõe o Sistema WEBMET. Com a implantação do Sistema WEBMET, o registro das informações meteorológicas passou a ser digital, e o armazenamento das informações impressas, obsoleto. Para manter um banco de dados coeso, a Força Aérea Brasileira (FAB) passou a digitalizar as informações até então registradas apenas em papel, inserindo-as também no banco de dados digital.

Os dados de altitude utilizados foram retirados do código TEMP, no formato *Unmerged*, disponível no sítio da Universidade de Wyoming. O código TEMP é gerado a partir de radiossondagens realizadas a cada 12 horas (0000

UTC¹ e 1200 UTC) pelas Estações Meteorológicas de Altitude. A principal função da radiossondagem é coletar informações de temperatura do ar, umidade relativa, direção e velocidade do vento, e pressão atmosférica entre a superfície e os níveis superiores da atmosfera. Estas informações são coletadas a partir de uma sonda composta de sensores, processadores, transmissores e antena GPS acoplada a um balão de gás.

A sondagem fornece dados do ar superior, os quais auxiliam na previsão de diversos fenômenos meteorológicos significativos para a aviação como turbulência, formação de gelo em aeronaves, probabilidade de trovoadas, formação de nuvens, trilhas de condensação, movimento e dispersão de cinzas vulcânicas e nuvens radioativas.

3.2.1 Tratamento dos dados

Os dados de superfície disponibilizados pelo ICEA estavam em ordem cronológica, identificados por data e hora, e separados por variável. Para esta pesquisa foram considerados os seguintes dados horários:

- a) hora;
- b) data;
- c) visibilidade horizontal predominante (vis);
- d) precipitação acumulada (precip);
- e) temperatura do ar (temp);
- f) temperatura do ponto de orvalho (PO);
- g) cobertura total de nuvens (nuvem);
- h) pressão atmosférica ao nível da estação (pressão);
- i) direção do vento (dir);
- j) velocidade do vento (vel); e
- k) umidade relativa do ar (UR).

Foram consideradas apenas as informações regulares. Os dados de informações especiais e locais foram ignorados

Para os dados de ar superior, foram coletadas os seguintes dados, com intervalo de 12 horas:

¹ Horário Internacional padrão. Para o meridiano em que se encontra a cidade de Curitiba, somam-se 3 horas ao horário brasileiro.

- a) velocidade do vento no nível de pressão de 250hPa (vel250);
- b) altura geopotencial do nível de pressão de 850 hPa (GPH850);
- c) altura geopotencial do nível de pressão de 500hPa (GPH500);
- d) temperatura do ar no nível de pressão de 700 hPa (temp700);
- e) temperatura do ponto de orvalho no nível de pressão de 700 hPa (PO700);
- f) temperatura do ar no nível de pressão de 500hPa (temp500); e
- g) temperatura do ponto de orvalho no nível de pressão de 500hPa (PO500).

Os níveis de pressão de 250hPa, 500hPa, 700hPa e 850hPa correspondem a, aproximadamente, 10Km, 5Km, 3Km e 1,5Km de altitude. Estes dados foram selecionados por permitirem a indicação da presença de fenômenos de escala sinótica, que podem interferir nos ocorridos à superfície.

O vento no nível de pressão de 250hPa pode indicar a ação de uma corrente de jato, o que pode significar a atuação de um sistema de baixa pressão, o que desfavorece a formação de nevoeiros devido ao movimento ascendente do ar; as alturas geopotenciais nos níveis de pressão de 850hPa e 500hPa podem indicar maior ou menor baroclinicidade atmosférica, sinalizando, por exemplo, a presença de centros de alta pressão, o que favorece a formação do fenômeno, devido ao comportamento descendente do ar; e as temperaturas do ar e temperaturas do ponto de orvalho nos níveis de pressão de 700hPa e 500hPa fornecem o gradiente de umidade na camada inferior da atmosfera.

3.2.1.1 Controle de qualidade dos dados

Os dados foram tratados visando apenas a exclusão de variáveis que indicassem discrepâncias, tais como:

- a) Temperaturas do ar inferiores a -20°C e superiores a 50°C;
- b) visibilidades horizontais predominantes inferiores a 0 metros, ou superiores a 20km (pois este é o registro máximo permitido no WEBMET);
- c) temperaturas do ponto de orvalho superiores à temperatura do ar, ou fora do intervalo entre -20°C e 50°C;

- d) velocidade do vento inferior a 0 nós ou superior a 50 nós (92,6 Km/h);
- e) direções do vento fora do intervalo de 0° a 360°; e
- f) umidades relativas inferiores a 10% ou superiores a 100%.

Para as variáveis que a princípio não têm um intervalo de validade bem definido, tais como a precipitação e a pressão atmosférica, foram analisados os máximos e mínimos de cada variável. Para o caso da precipitação, o valor mínimo registrado na amostra foi 0mm, e o valor máximo foi 56,7mm, valores que são possíveis de ocorrer. Para os dados de pressão, o valor mínimo registrado na amostra foi 900hPa, e o máximo foi 951,6hPa, que são também plausíveis.

A variável nebulosidade, como já citado, é informada em oitavos, de 0 a 8. No entanto, quando por uma restrição de visibilidade não é possível enxergar a cobertura de nuvens, o valor informado como cobertura de nuvens é 9. Quando nesta situação, este valor foi substituído pela última nebulosidade observada antes da restrição. Após a substituição foram retirados valores inferiores a zero, ou superiores a 8 oitavos.

Os dados detectados como discrepantes foram substituídos por “9999”, pois este valor não representa confusão com nenhuma medida real. Esta medida foi adotada, pois a simples exclusão da informação não era possível devido à perda da sequência temporal em relação às demais variáveis. Os dados de altitude não foram tratados.

3.2.1.2 Pré-processamento dos dados

3.2.1.2.1 Pré-processamento das informações de direção do vento

A variável direção do vento é uma variável cíclica, e por este motivo não determina uma quantidade em si, apenas uma orientação. Por esta razão foi necessária a polarização deste dado, de modo que a rede neural não hierarquizasse os valores. A direção do vento foi dividida em quatro grupos, de acordo com os quadrantes, designados por sequências de números 1 e -1, onde três dos valores são -1, e o valor 1, que ativa o neurônio cuja posição refere-se ao intervalo em que está a referida direção do vento. Assim, direções

do vento compreendidas no primeiro intervalo, entre 0° e 90°, foram designadas pela sequência 1 -1 -1 -1; direções do vento compreendidas no segundo intervalo, entre 90° e 180°, foram designado pela sequência -1 1 -1 -1; direções do vento compreendidas no terceiro intervalo, entre 180° e 270°, foram designadas pela sequência -1 -1 1 -1; e direções do vento compreendidas no quarto intervalo, entre 270° e 360°, foram designadas pela sequência -1 -1 -1 1, conforme apresentado na Tabela 3, abaixo.

TABELA 3- POLARIZAÇÃO DA VARIÁVEL DIREÇÃO DO VENTO

Direção do vento (em graus)	Sequência designadora
[000,090)	1 -1 -1 -1
[090,180)	-1 1 -1 -1
[180,270)	-1 -1 1 -1
[270,360)	-1 -1 -1 1

FONTE: A Autora, 2016.

3.2.1.2.2 Pré-processamento das informações de visibilidade

Como esta pesquisa trata da previsão categórica do fenômeno nevoeiro, as informações de visibilidade foram utilizadas não apenas em seu formato métrico, como também em seu formato categórico. Os casos em que se verificaram visibilidades inferiores a 1000m, e, por conseguinte, a presença do fenômeno, foram tomados como 1; e os casos em que a visibilidade registrada foi igual, ou superior, a 1000m, e, portanto, não configuraram ocorrência de nevoeiro, foram tomados como -1.

Esta medida foi adotada no intuito de auxiliar o trabalho da Rede Neural Artificial, pois simplifica a representação da variável em questão, atribuindo-lhe apenas dois valores possíveis: 1 e -1.

3.2.1.3 Compatibilização da amostra dos dados de altitude

Diferentemente das observações à superfície, as radiossondagens não ocorrem de hora em hora, e sim a cada 12 horas. Para que o número de informações de altitude fosse compatível com o número de informações de superfície, foi realizado o complemento da amostra dos dados sinóticos. Para

tanto, cada informação de altitude foi repetida onze vezes, nas onze horas subsequentes ao horário efetivo da sondagem, mantendo, assim, a compatibilidade da série temporal entre as informações de superfície e altitude.

O procedimento foi adotado considerando que a escala temporal de variação das grandezas sinóticas (ordem 10^5 s) é suficientemente maior do que a variações locais das condições em superfície (ordem 10^3 s). Dessa forma, as eventuais alterações nas variações de ar superior não têm impacto significativo na visibilidade em curto prazo.

3.2.2 Base de dados tratada

Ao final do tratamento dos dados, a amostra pode contar com 157683 informações de visibilidade horizontal predominante (vis), sendo 7300 destas, de visibilidades inferiores a 1000 metros, 155495 informações de precipitação acumulada (precip), 154733 informações de temperatura do ar (temp), 154733 informações de temperatura do ponto de orvalho (PO), 145145 informações de cobertura total de nuvens (nuvem), 153265 informações de pressão atmosférica ao nível da estação (pressão), 115820 informações de direção do vento (dir), 115820 informações de velocidade do vento (vel), 154337 informações de umidade relativa do ar (UR), 94440 informações de velocidade do vento no nível de pressão de 250hPa (vel250), 109776 informações de altura geopotencial do nível de pressão de 850hPa (GPH850), 108504 informações de altura geopotencial do nível de pressão de 500hPa (GPH500), 109836 informações de umidade relativa no nível de pressão de 700hPa (UR700), e 109704 informações de umidade relativa no nível de pressão de 500hPa (UR500), totalizando 1839292 registros, conforme é possível observar na Tabela 4, a seguir.

TABELA 4- NÚMERO DE REGISTROS DA AMOSTRA DE DADOS

Variável	Número de registros
vis	157683
precip	155495
temp	154733
PO	154733
nuvem	145145
pressão	153265
dir	115820
vel	115820
UR	154337
vel250	94440
GPH850	109776
GPH500	108504
UR700	109836
UR500	109704
Total	1839292

FONTE: A Autora, 2016.

Como o código TEMP não fornece, diretamente, dados de umidade relativa, trata-se, nesta pesquisa, como umidade relativa nos níveis de pressão de 700hPa e 500hPa, como a apresentação simultânea à RNA, dos dados de temperatura do ar e ponto de orvalho nos respectivos níveis de pressão.

3.3 MÉTRICAS PARA AVALIAÇÃO DE DESEMPENHO DE PREVISÕES DE VARIÁVEIS CATEGÓRICAS

Para a avaliação do desempenho das previsões citadas nesta pesquisa foram utilizados os métodos descritos por Wilks (2006).

Como já citado, a previsão do fenômeno nevoeiro neste trabalho será tratada apenas em sua forma categórica. Ou seja, serão consideradas apenas duas classes de ocorrência do evento: ou o evento ocorreu, e a esta situação chamaremos classe; ou o evento não ocorreu, e a esta situação chamaremos não-classe.

Em relação à previsibilidade, um evento pode ser previsto ou não previsto. Em relação à ocorrência, ele pode ser observado ou não observado. Logo, dado um fenômeno qualquer, este certamente encontra-se em uma das quatro combinações possíveis de previsão e ocorrência. A saber,

- a) Foi previsto e observado (P-O);
- b) Foi previsto, mas não observado (P-NO);
- c) Não foi previsto, mas foi observado (NP-O);
- d) Não foi previsto e não foi observado (NP-NO).

Estas possibilidades podem ser representadas, de forma esquemática, em uma tabela de contingência, conforme a Tabela 5, abaixo.

TABELA 5- TABELA DE CONTINGÊNCIA

		Observado		
		Classe	Não-Classe	Total
Previsto	Classe	a	b	a+b
	Não-Classe	c	d	c+d
	Total	a+c	b+d	n=a+b+c+d

FONTE: Adaptado de Wilks, 2006.

onde *a* representa o número de casos previstos e observados; *b* representa os casos em que o evento foi previsto, mas não observado; *c* representa o número casos não previstos, mas observados; *d* indica os casos não previstos e não observados; e *n* indica a soma total do número de casos.

3.3.1 Medidas escalares de desempenho

O uso de medidas escalares na caracterização do desempenho de previsões facilita a compreensão dos resultados alcançados se comparado à utilização de frequências relativas ou absolutas (WILKS, 2006). Destacam-se, abaixo, algumas das medidas escalares mais utilizadas na verificação de desempenho de previsões.

3.3.1.1 Proporção Correta (PC)

Esta métrica avalia a proporção de acertos da previsão, tanto para os casos P-O quanto para os casos NP-NO. É simplesmente, a fração entre o número de acertos da previsão, e os casos totais.

$$PC = \frac{a + d}{n} \quad (7)$$

Wilks (2006) salienta que a Proporção Correta satisfaz o princípio da equivalência de eventos, visto que o valor atribuído à ocorrência prevista de um caso é o mesmo atribuído a não-ocorrência de um evento não-previsto.

No entanto, esta métrica pode não ser bem aplicada em casos de eventos raros, pois a não-ocorrência do evento é muito mais provável que a ocorrência, o que torna estes casos bem mais difíceis de serem previstos. Além disso, a penalidade para os erros da previsão é a mesma para os casos de P-NO e NP-O.

3.3.1.2 Índice de Sucesso Crítico (*Threat Score*- ISC)

Esta métrica é bem mais ajustada a casos de eventos raros, pois desconsidera os casos de não-evento, mais abundantes e de maior previsibilidade.

O *ISC* consiste na fração obtida pelo número de casos em que o fenômeno foi previsto e de fato ocorreu, e o total de casos que envolvem o fenômeno raro (o fenômeno foi previsto e/ou observado).

$$ISC = \frac{a}{a + b + c} \quad (8)$$

3.3.1.3 Viés (V)

O Viés fornece a informação de quantas vezes o evento foi previsto, em relação a quantas vezes foi observado, sem levar em consideração a correspondência entre observações e previsões. Por isso não é considerado uma medida de precisão.

$$V = \frac{a + b}{a + c} \quad (9)$$

Valores de V maiores que 1, indicam uma previsão superestimada do evento, ou seja, o evento foi previsto mais do que ocorreu; $V=1$ indica uma previsão imparcial, ou seja, o evento foi previsto o mesmo número de vezes que foi observado; valores de V menores que 1 indicam uma previsão subestimada do evento, onde o número de previsões é inferior ao número de ocorrências.

3.3.1.4 Razão de Falso Alarme (RFA)

A RFA representa a proporção de previsões que acabaram por não se concretizar. É dada pelo número de casos P-NO dividido pelo número total de vezes que o evento foi observado.

$$RFA = \frac{b}{a + b} \quad (10)$$

Esta métrica assume valores entre 0 e 1. No entanto, como se refere a uma medida com significado desfavorável, pois indica erros, quanto mais próxima de zero, mais eficaz é a ferramenta de previsão.

3.3.1.5 Probabilidade de Detecção (*Hit Rate*- PD)

É, basicamente, a taxa de acerto da previsão relacionada apenas ao evento de interesse. É dada pela razão entre o número de casos P-O e o número total de ocorrências do evento.

$$PD = \frac{a}{a + c} \quad (11)$$

Operacionalmente, é de suma importância que uma ferramenta de previsão seja eficaz em dois aspectos principais: capacidade de prever corretamente a ocorrência do fenômeno de interesse; e manter alta a confiabilidade das previsões, não emitindo falsos alarmes em demasia. Dada a motivação operacional desta pesquisa, as métricas de avaliação de

desempenho da Rede Neural Artificial terão a seguinte hierarquia: como métrica mais importante, a que será primeiramente observada será a Probabilidade de Detecção (PD), pois este índice caracteriza bem o índice de acertos da ferramenta em relação aos casos nos quais o evento de fato ocorreu. Como segunda métrica, será utilizada a Razão de Falso Alarme (RFA), visto que é de interesse operacional não restringir as condições meteorológicas desnecessariamente. A RFA é um índice importantíssimo operacionalmente, pois é ela que garante a manutenção da confiabilidade dos usuários em relação às previsões fornecidas. Se por absurdo, a condição de nevoeiro fosse prevista todos os dias, para todos os horários, o índice de acerto principal desta pesquisa seria 100%, pois todos os eventos ocorridos, de fato, foram previstos. No entanto, o grande número de falsos alarmes emitidos faria as previsões caírem em descrédito, visto que condição não se estabelece na mesma medida em que é prevista. Como terceira métrica será utilizado o Índice de Sucesso Crítico (ISC), pois, segundo Wilks (2006), este índice destaca a previsibilidade de fenômenos raros, o que é o caso do fenômeno nevoeiro. Como quarta, e última, medida será considerada a Proporção Correta (PC), pois para casos de fenômenos raros esta é a menos significativa. Mais uma vez se, por absurdo, nunca fosse prevista a condição de nevoeiro, este índice seria bastante alto, pois os casos de não-ocorrência, muito mais numerosos, sempre seriam previstos corretamente, produzindo assim mesmo um alto valor de PC.

A ordem hierárquica estabelecida pode ser observada, de forma esquemática na Tabela 6, abaixo.

TABELA 6- HIERARQUIA ENTRE AS MÉTRICAS

1ª- Probabilidade de Detecção- PD
2ª- Razão de Falso Alarme- RFA
3ª- Índice de Sucesso Crítico- ISC
4ª- Proporção Correta- PC

FONTE: A Autora, 2016.

3.4 CLIMATOLOGIA E PERSISTÊNCIA

Como comparativo aos resultados que serão obtidos a partir do emprego de RNAs RBF na previsão do fenômeno nevoeiro, foram realizadas outras duas análises com métodos de previsão sem destreza (*no skill*): climatologia e persistência. Mais detalhes sobre estes estudos serão apresentados a seguir.

3.4.1 Climatologia

Uma das maneiras de buscar entender melhor um fenômeno meteorológico é estudar sua climatologia. Nesta seção será apresentada a climatologia de diversas variáveis meteorológicas associadas aos eventos de nevoeiro no Aeroporto Afonso Pena.

3.4.1.1 Distribuição dos eventos por mês e horário

Em relação ao fenômeno nevoeiro, podemos verificar a frequência de sua distribuição nos meses do ano. Obteve-se a distribuição apresentada na Tabela 7, considerando o número de horas em que houve nevoeiro, no período de 1997 a 2014:

TABELA 7- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NOS MESES DO ANO

Meses	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
Número de ocorrências	228	210	408	605	1079	1336	1128	929	551	366	250	210

FONTE: A Autora, 2016.

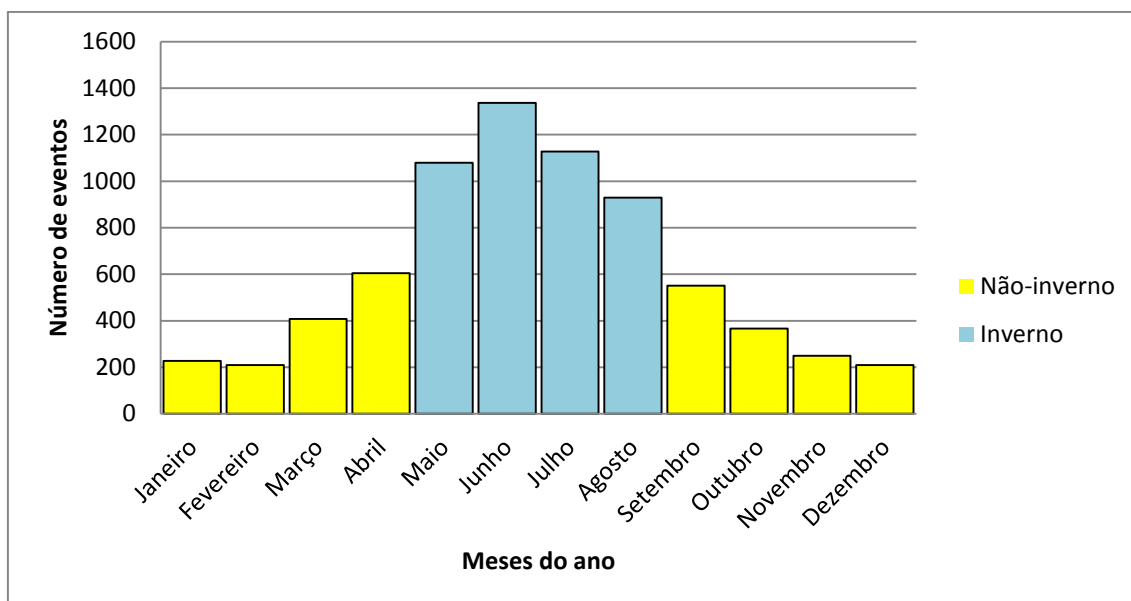


FIGURA 8- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NOS MESES DO ANO
FONTE: A Autora, 2016.

Na Figura 8, é facilmente perceptível a semelhança da distribuição do fenômeno durante os meses do ano, e uma curva gaussiana. Nota-se que o número de ocorrências de nevoeiro é maior quanto mais próximo do inverno está o mês em questão, tendo sua maior concentração no mês de junho. O período que se destaca em relação à ocorrência de nevoeiros é o que compreende os meses de maio a agosto, período que concentra, aproximadamente, 61% dos casos de nevoeiro. Tomaremos este período por inverno, e os demais meses do ano como não-inverno.

Em relação à ocorrência de nevoeiro nas horas do dia (em horário local), temos a distribuição apresentada na Tabela 8 e a respectiva ilustração na Figura 9.

TABELA 8- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NAS HORAS DO DIA

Número de ocorrências	Hora											
	0	1	2	3	4	5	6	7	8	9	10	11
	391	472	559	622	696	776	875	894	566	288	86	20
	12	13	14	15	16	17	18	19	20	21	22	23
	21	31	36	41	46	48	53	47	71	136	210	315

FONTE: A Autora, 2016.

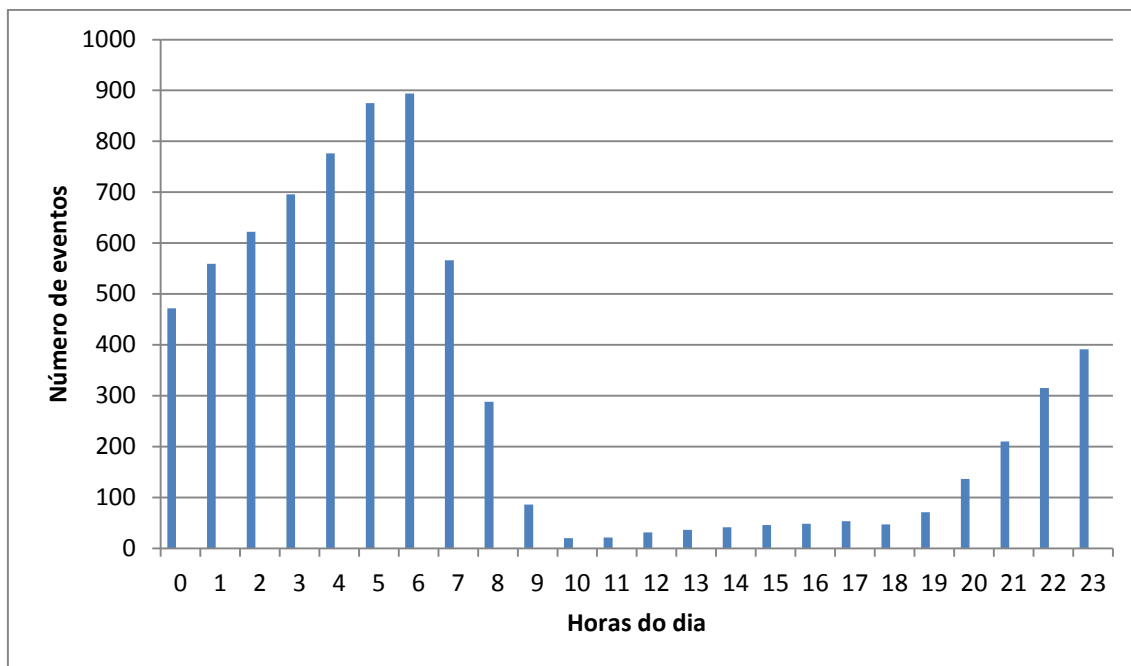


FIGURA 9- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO NAS HORAS DO DIA
 FONTE: A Autora, 2016.

É possível observar que o período mais propício à ocorrência de nevoeiros está no intervalo entre 21 e 8 horas local. Nota-se também que há um aumento no número de ocorrências do fenômeno no mesmo período. Com o pôr do sol inicia-se um processo acumulativo de perda de radiação. O solo torna-se então cada vez mais frio, favorecendo a gênese do fenômeno. A partir das 7 horas local, nota-se um decréscimo no número de ocorrências de nevoeiro. Esta condição se deve ao fato de que, com o nascer do sol a superfície volta a receber radiação, torna a aquecer, e a condição favorável à manutenção do fenômeno se extingue. Nota-se que este processo ocorre de forma gradual, pois da mesma forma que a temperatura do solo decresce gradativamente devido à perda acumulativa de radiação, seu aquecimento ocorre também de forma gradativa, a partir de um ganho acumulativo de radiação.

3.4.1.2 Distribuição dos eventos por faixas de temperatura do ar

Tomando intervalos de temperatura do ar de cinco em cinco graus Celsius, obtém-se a seguinte distribuição para o momento de ocorrência do fenômeno, e 1, 2 e 3 horas antes de sua configuração, denotadas por “0h”,

“-1h”, “-2h” e “-3h”, respectivamente, conforme apresentado na Tabela 9, e ilustrado na Figura 10.

TABELA 9- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A TEMPERATURA DO AR

	Temperatura (em °C)	[-5,0)	[0,5)	[5,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,35)	[35,40)
Número de ocorrências	0h	12	250	1361	3217	1912	278	23	0	0
	-1h	11	1336	3231	1918	249	71	5	0	0
	-2h	11	1253	3193	2013	273	89	14	0	0
	-3h	10	1133	3139	2132	351	97	19	0	0

FONTE: A Autora, 2016.

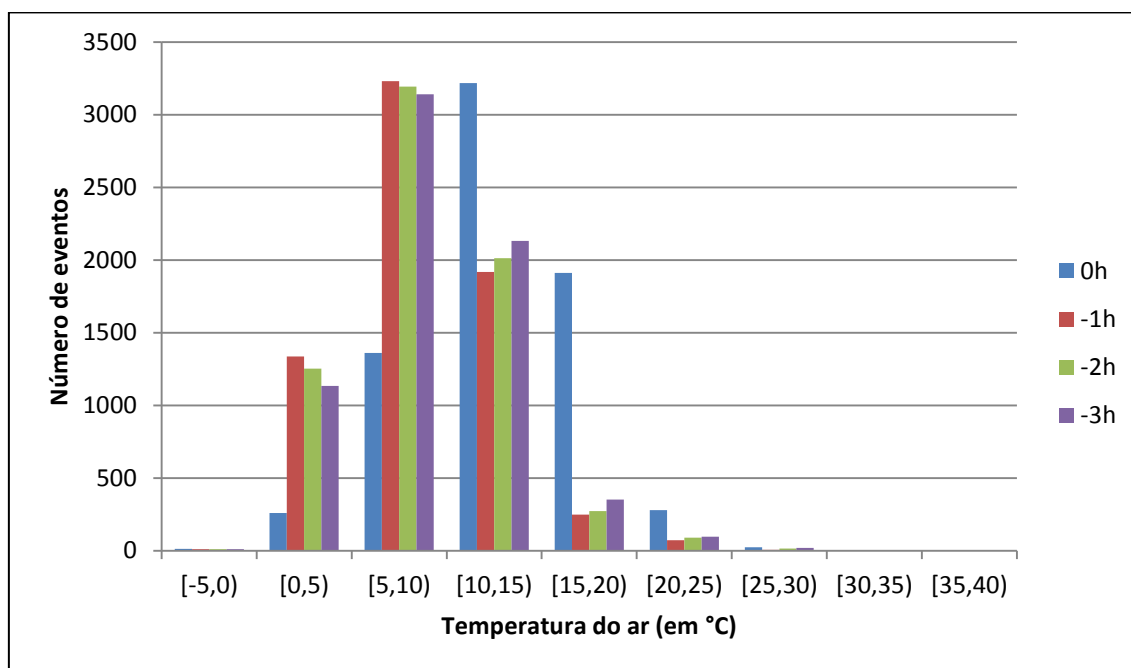


FIGURA 10- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A TEMPERATURA DO AR

FONTE: A Autora, 2016.

Nota-se que a maior ocorrência do fenômeno dá-se quando a temperatura do ar está entre 15 e 20°C, e que nas três horas que antecedem o fenômeno a temperatura tende a estar mais baixa, entre 10 e 15°C.

É possível observar que as temperaturas em que mais ocorre a formação de nevoeiro são justamente as temperaturas típicas do período que

compreende as primeiras horas da noite e se estende até as primeiras horas da manhã.

Estas informações de temperatura e período do dia de maior ocorrência de nevoeiro vêm ao encontro do que já se conhece do fenômeno. Os períodos do dia de maior ocorrência são justamente aqueles em que o ar imediatamente superior ao solo está menos turbulento devido ao aquecimento solar.

3.4.1.3 Distribuição dos eventos por faixas de umidade relativa

Observando o comportamento da variável umidade relativa quando na ocorrência do fenômeno nevoeiro, obtém-se o apresentado na Tabela 10, e ilustrado na Figura 11, que a segue.

TABELA 10- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A UMIDADE RELATIVA

		Umidade Relativa (em %)									
		[0,10)	[10,20)	[20,30)	[30,40)	[40,50)	[50,60)	[60,70)	[70,80)	[80,90)	[90,100]
Número de ocorrências por horário	0h	0	0	4	1	13	12	27	41	111	6702
	-1h	0	0	4	7	13	24	49	60	123	6669
	-2h	0	0	5	9	19	43	58	81	167	6584
	-3h	0	0	5	16	33	49	91	116	263	6415

FONTE: A Autora, 2016

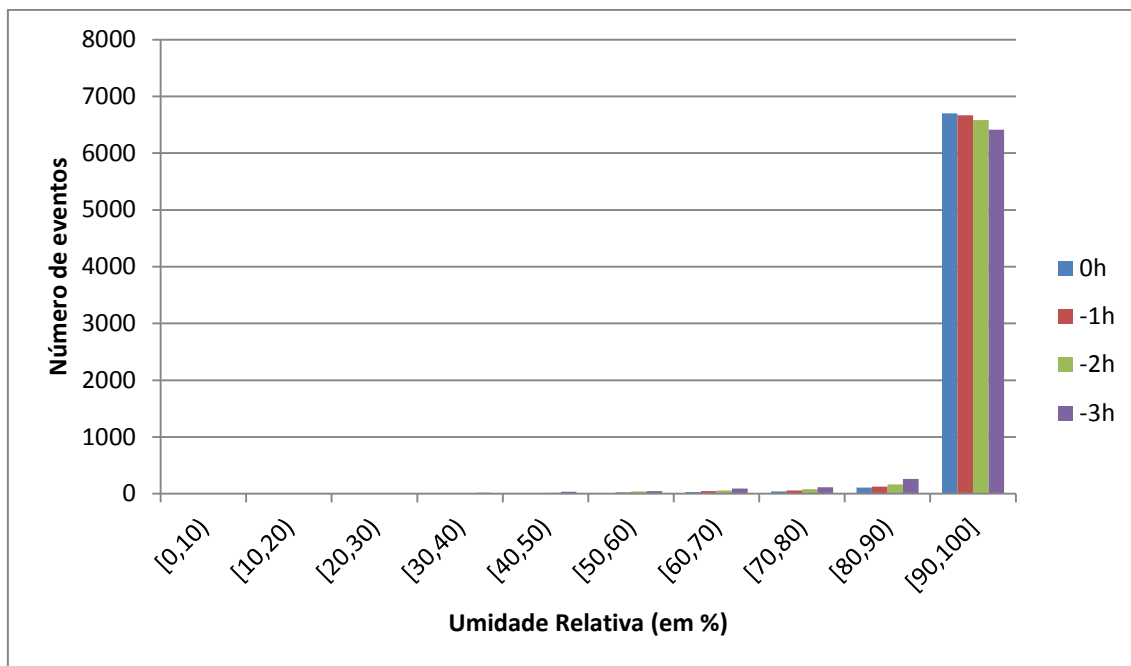


FIGURA 11- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A UMIDADE RELATIVA
 FONTE: A Autora, 2016

Fica clara a forte relação entre a condição de umidade relativa acima dos 90% e a ocorrência do fenômeno nevoeiro. Nota-se que existem casos em que houve restrição severa de visibilidade, mas a umidade relativa se manteve bastante baixa. É importante destacar que existem outros fenômenos restritores de visibilidade como poeira, areia ou fumaça. Nestes casos, o ar não necessita estar próximo à saturação, tal como no nevoeiro. Assim, os casos que podem ser observados na Tabela 10 e na Figura 9, em que a visibilidade se restringiu a menos de 1000 metros, mas a umidade relativa se manteve bastante baixa podem ser atribuídos a estes outros fenômenos.

3.4.1.4 Distribuição dos eventos em relação à cobertura de nuvens

Em relação à cobertura de nuvens nas três horas que antecedem o fenômeno, bem como no momento de sua ocorrência, obtém-se o disposto na Tabela 11 e o ilustrado na Figura 12.

TABELA 11- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A COBERTURA DE NUVENS

	Cobertura de nuvens (em oitavos)	0	1	2	3	4	5	6	7	8
Número de ocorrências por horário	0h	46	194	277	728	535	919	581	1332	2205
	-1h	57	253	396	813	583	913	532	1224	1929
	-2h	49	342	468	815	577	892	560	1152	1655
	-3h	42	428	522	809	539	906	564	1118	1448

FONTE: A Autora, 2016

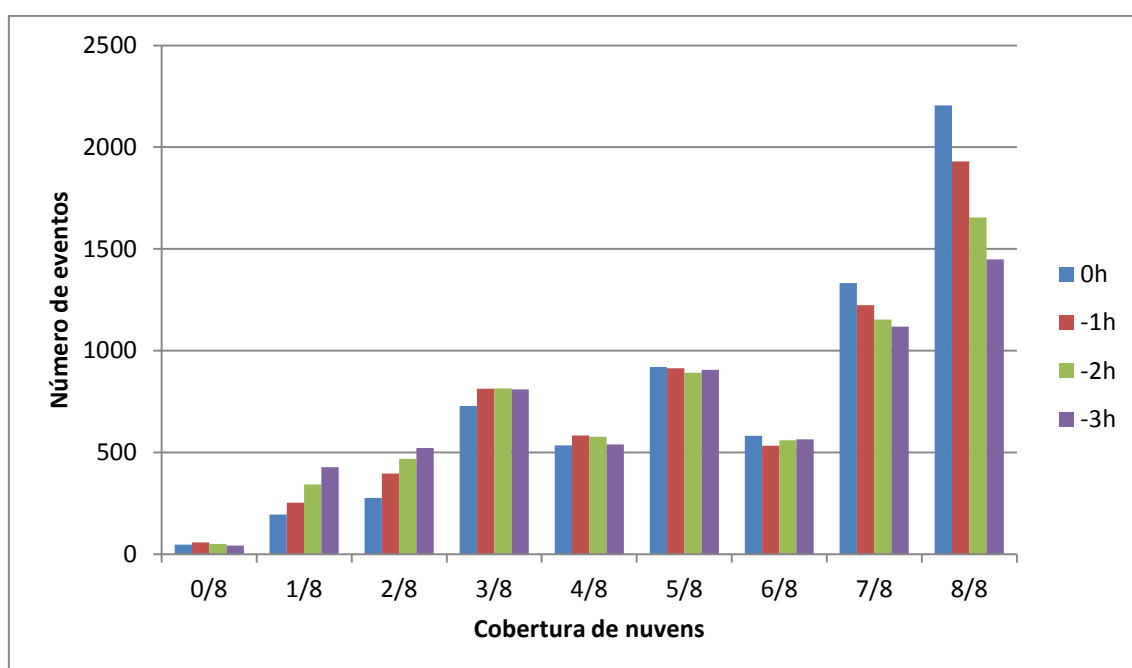


FIGURA 12- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A COBERTURA DE NUVENS

FONTE: A Autora, 2016

Nota-se que a cobertura de nuvens, quando na ocorrência de nevoeiros, é de céu nublado, ou completamente encoberto, predominando as coberturas de 7 e 8 oitavos.

É importante lembrar que foi tomada a cobertura total de nuvens no céu, independentemente do tipo ou altura da camada. Por esta razão, talvez, não se tenha observado muitas ocorrências de nevoeiros em céu claro, pois o observador deve informar, no formulário WEBMET, qualquer nebulosidade nas vizinhanças do aeródromo, por menor ou mais alta que seja.

3.4.1.5 Distribuição dos eventos em relação à pressão atmosférica

Em relação à pressão atmosférica reinante na estação nas três horas antecedentes ao fenômeno, e na hora de sua ocorrência temos o disposto na Tabela 12 e na Figura 13 que seguem.

TABELA 12- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRESSÃO ATMOSFÉRICA

		Pressão atmosférica (em hPa)						
		[900,905)	[905,910)	[910,915)	[915,920)	[920,925)	[925,930)	[930,935)
Número de ocorrências por horário	0h	7	408	2321	3750	583	1	0
	-1h	9	424	2344	3750	543	0	0
	-2h	9	412	2350	3781	519	0	0
	-3h	8	397	2330	3825	512	0	0

FONTE: A Autora, 2016

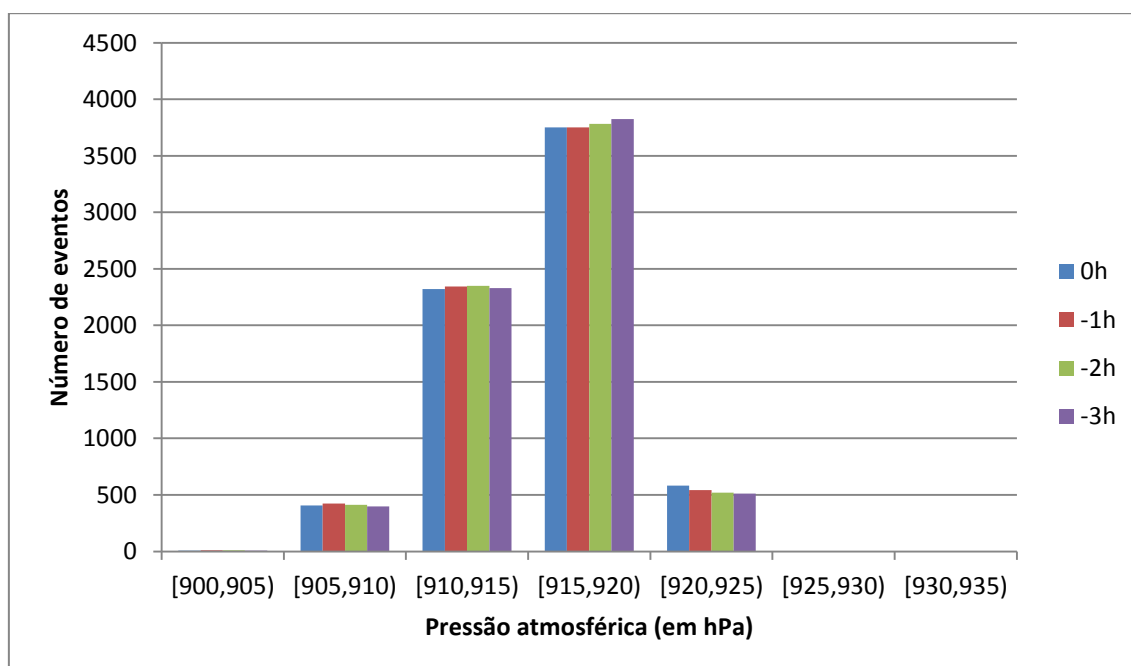


FIGURA 13- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRESSÃO ATMOSFÉRICA

FONTE: A Autora, 2016

Nota-se que a maior parte das ocorrências de nevoeiro se dá quando a pressão atmosférica ao nível da estação está compreendida no intervalo entre 915 e 920hPa.

3.4.1.6 Distribuição dos eventos em relação à direção e à velocidade do vento

As variáveis direção e velocidade do vento são de grande importância no estudo dos nevoeiros. Tanto que, para um tipo específico do fenômeno (nevoeiro de advecção), as informações de vento são condicionantes da ocorrência do evento.

Observando a direção predominante do vento, no momento de ocorrência do fenômeno nevoeiro, em intervalos de 10 em 10 graus, é possível distribuir o número de eventos conforme o disposto na Tabela 13, onde a direção 0° indica o norte verdadeiro, a direção 90° indica o leste, a direção 180° indica o sul, e a direção 270° indica o oeste.

TABELA 13- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO

Direção do vento (em graus)	Número de casos							
	0	1592	90	282	180	51	270	41
	10	79	100	292	190	34	280	18
	20	173	110	241	200	37	290	23
	30	268	120	285	210	43	300	38
	40	307	130	167	220	34	310	52
	50	328	140	122	230	34	320	56
	60	402	150	85	240	42	330	88
	70	283	160	53	250	20	340	79
	80	284	170	54	260	26	350	118
Total	[0,80]	3716	[90,170]	1581	[180,260]	321	[270,350]	513

FONTE: A Autora, 2016

Nota-se que uma boa forma de tomar intervalos de tamanhos maiores do que 10 graus, mantendo uma hierarquização entre os grupos, é a divisão em quadrantes. Assim, fazendo o levantamento para o vento em quadrantes, nas três horas que antecedem o fenômeno, e no momento do evento, obtém-se a distribuição apresentada na Tabela 14 e ilustrada na Figura 14 seguintes.

TABELA 14- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO EM QUADRANTES

	Direção do vento (em graus)	[0,90)	[90,180)	[180,270)	[270,360)
Número de ocorrências por horário	0h	3323	1298	284	373
	-1h	3270	1451	238	314
	-2h	3111	1646	217	296
	-3h	2908	1862	210	292

FONTE: A Autora, 2016

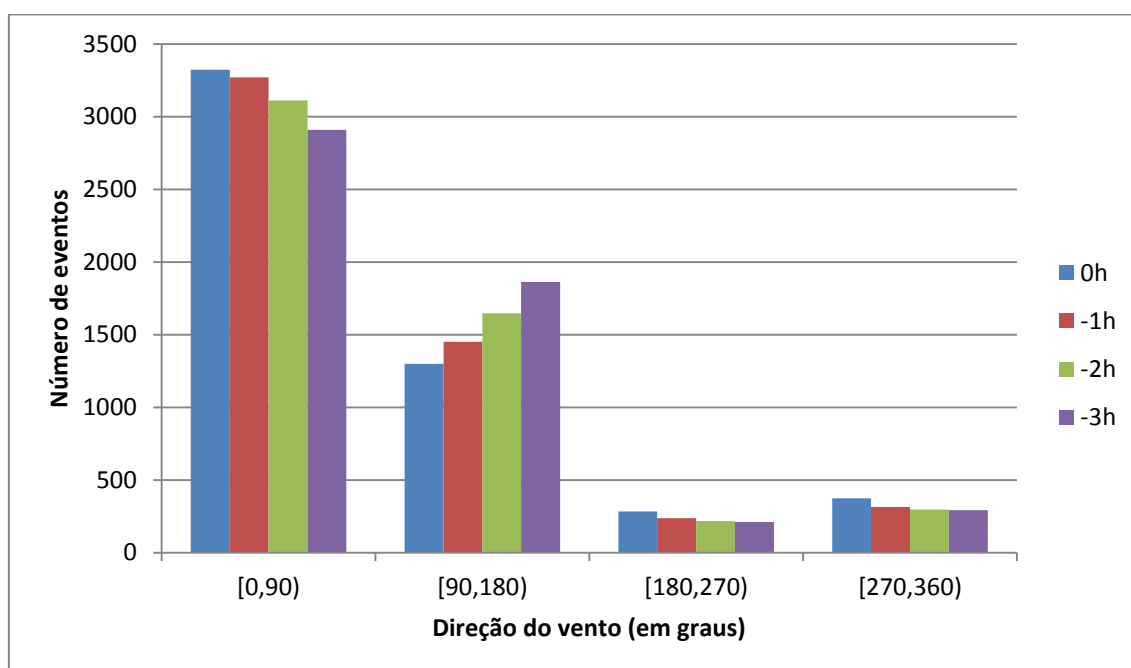


FIGURA 14- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A DIREÇÃO DO VENTO EM QUADRANTES

FONTE: A Autora, 2016

A grande maioria dos casos de nevoeiro ocorre com vento soprando de direções contidas no intervalo entre 0 e 90°. Como pode ser observado na Figura 7, a proximidade com a região oceânica permitiria que ventos soprando de qualquer direção compreendida entre 0 e 180° transportassem umidade do oceano. No entanto, a cadeia de montanhas a sudeste da cidade de São José dos Pinhais, talvez, impeça que ventos soprando de direções entre 90 e 180° acabem por configurar condições de nevoeiro no Aeroporto Afonso Pena. Isso, porque ventos quentes e úmidos soprando deste quadrante, ao defrontarem-se com a cadeia montanhosa, são forçados a subir e, resfriando-se adiabaticamente, condensam-se na encosta da própria montanha,

configurando a formação de um nevoeiro orográfico; ou ao perder o contato com o solo, configurem a formação de nuvens do tipo estratiformes. Logo, as características da porção de ar litorânea, que seriam favoráveis à formação do fenômeno no Aeroporto Afonso Pena, se perdem no caminho entre o oceano e a cidade, resultando em um número bastante inferior de casos de nevoeiro com ventos soprando da direção sudeste.

Já os ventos que sopram da direção nordeste, no intervalo entre 0 e 90°, têm menos obstáculos em seu trajeto. Por isso podem levar as porções de ar quentes e úmidas, da região oceânica à cidade de São José dos Pinhais, com menores perdas destas características. Esta relação pode indicar o tipo de nevoeiro característico do Aeroporto Afonso Pena como sendo de advecção.

Em relação a velocidade do vento, temos a distribuição dos eventos de nevoeiro apresentada na Tabela 15, e ilustrada na Figura 15, a seguir:

TABELA 15- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A VELOCIDADE DO VENTO

	Velocidade do vento (em nós)	[0,5)	[5,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,35)
Número de ocorrências por horário	0h	4426	736	84	23	8	0	1
	-1h	4438	705	112	17	1	0	0
	-2h	4356	793	102	18	1	0	0
	-3h	4259	887	110	14	2	0	0

FONTE: A Autora, 2016

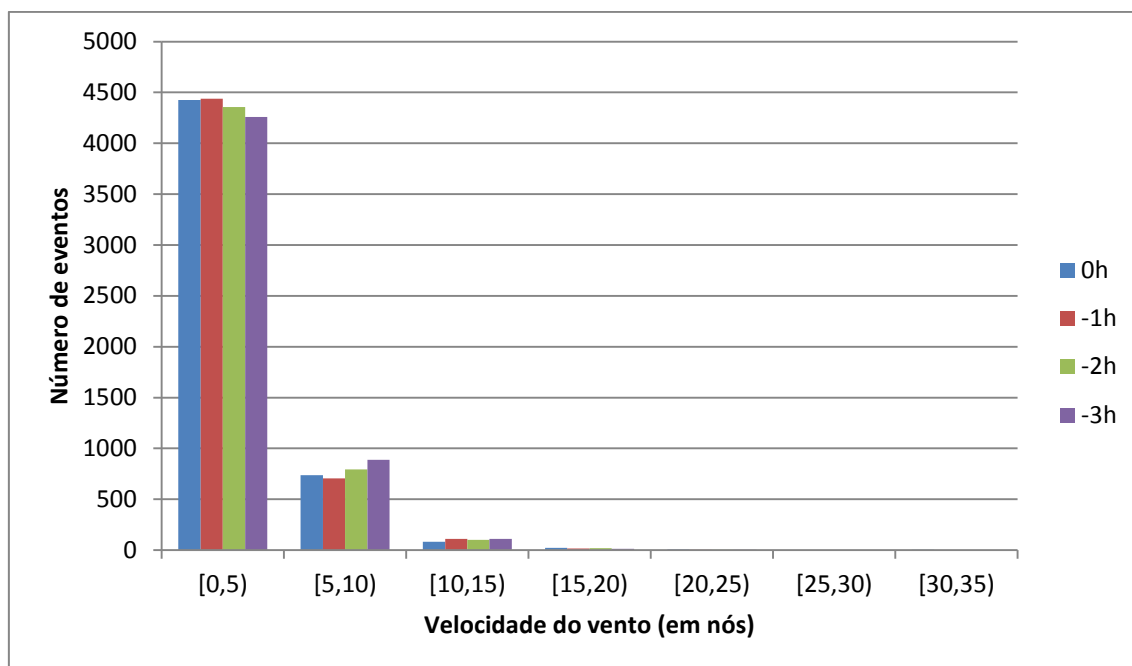


FIGURA 15- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A VELOCIDADE DO VENTO
FONTE: A Autora, 2016

O gráfico ilustra que a maior concentração de casos de nevoeiro ocorre quando a velocidade do vento está fraca, entre 0 e 5 nós. Esta condição reforça a constatação obtida em 3.4.1.6 de que a maioria dos casos de nevoeiro do Aeroporto Afonso Pena são do tipo de advecção, haja vista a definição deste fenômeno já citada na Seção 2.1.2.

3.4.1.7 Distribuição dos eventos em relação à precipitação

Distribuindo-se as ocorrências de nevoeiro em relação à ocorrência ou não-ocorrência de precipitação, obtém-se o apresentado na Tabela 16, cujos dados são ilustrados na Figura 16 que a segue.

TABELA 16- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRECIPITAÇÃO ACUMULADA

	Precipitação acumulada (em mm)	0	0-50
Número de ocorrências por horário	0h	7098	68
	-1h	7152	77
	-2h	7148	67
	-3h	7123	79

FONTE: A Autora, 2016

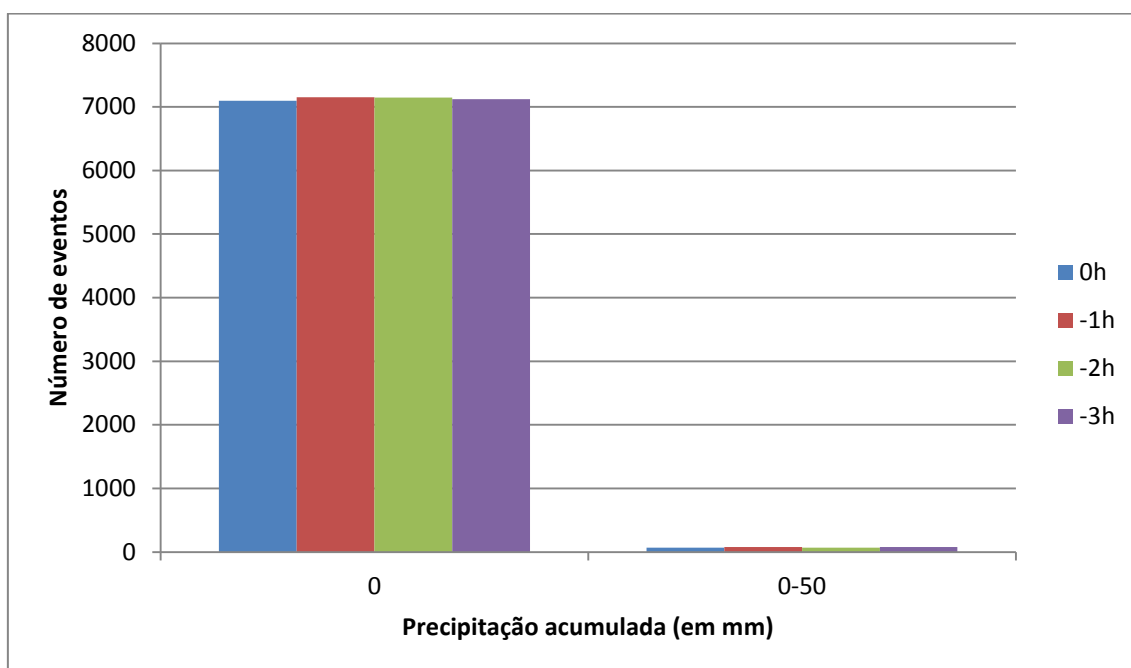


FIGURA 16- DISTRIBUIÇÃO DOS EVENTOS DE NEVOEIRO DE ACORDO COM A PRECIPITAÇÃO ACUMULADA

FONTE: A Autora, 2016

Fica claro, através da ilustração, que a maior parte dos casos de nevoeiro acontece quando não houve precipitação nem nas três horas anteriores, nem na hora de ocorrência do fenômeno.

Ocorrências de precipitação superiores a 50mm foram desconsideradas, pois, nestes casos, a restrição de visibilidade é causada pela precipitação, e não pelo fenômeno em questão nesta pesquisa.

3.4.1.8 Definição das previsões através de climatologia

A partir dos estudos realizados anteriormente, podemos observar o comportamento das variáveis de superfície nas horas que antecedem o evento nevoeiro. Através dessas análises, é possível estabelecer um modelo conceitual qualitativo de classificação de ocorrência (classe= 1) ou não-ocorrência (classe= -1) futura do fenômeno, sem especificar a probabilidade de ocorrência do evento. As condições favoráveis são as seguintes:

- Os meses mais favoráveis são maio, junho, julho e agosto;
- o horário de maior ocorrência é entre 21 e 8 horas local;
- a temperatura do ar deve estar compreendida entre 10 e 15°C;

- d) a umidade relativa deve ser superior a 90%;
- e) o céu deve estar nublado ou encoberto, com cobertura de nuvens igual a 7 ou 8 oitavos;
- f) a pressão atmosférica deve estar no intervalo entre 915 e 920hPa;
- g) o vento deve estar soprando de nordeste, entre 0 e 90°, e com pouca intensidade, entre 0 e 5 nós; e
- h) não deve ter havido precipitação nas três horas que precedem o fenômeno.

Esquemáticamente, estas condições podem ser ilustradas na tabela abaixo, onde as condições favoráveis à formação de nevoeiro foram atribuídas à classe (1), e as condições não favoráveis foram atribuídas à não-classe (-1).

TABELA 17- CONDIÇÕES FAVORÁVEIS À FORMAÇÃO DE NEVOEIRO

	Parâmetros da classificação	1	-1
Variável	mês	[5,8]	[1,5)U(8,12]
	hora	[21,8]	-8,21
	Temp (°C)	[10,15)	[-20,10)U[15,50)
	UR (%)	≥ 90	< 90
	Nuvem (oitavos)	[7,8]	[0,7)
	Pressão (hPa)	[915,920)	[900,915)U[920,952)
	Vel (nós)	[0,5)	≥ 5
	Precip (mm)	0	≠ 0

FONTE: A Autora, 2016.

É possível realizar previsões através destas informações, inferindo que, na ocorrência concomitante destas condições, haverá a formação de nevoeiro. Este experimento é realizado posteriormente, na Seção 4.1.

3.4.2 Persistência

Condições meteorológicas não são estanques, pois os fenômenos que acontecem na atmosfera dependem das condições reinantes, e estas, por sua vez, dependem de condições passadas. Pode-se dizer que a meteorologia acontece de forma contínua, onde a consumação de um fenômeno guarda relações com as condições anteriores, e sua ocorrência interfere nas condições posteriores. Dado isso, pode-se dizer que as variáveis atmosféricas têm uma

dependência estatística entre seus valores passado, presente e futuro (WILKS, 2006).

A esta dependência entre condição e tempo, em meteorologia, dá-se o nome de Dependência de Série Positiva, ou Persistência. Ou seja, um fenômeno é persistente se seus valores anteriores e posteriores são estatisticamente dependentes. Uma forte relação de dependência, ou uma forte persistência, permite fazer inferências estatísticas tão melhores quanto maior for esta relação.

Fenômenos constatados persistentes possuem uma técnica de previsão bastante simples: a condição futura é inferida como a manutenção da condição atual. Assim, um questionamento interessante a respeito do fenômeno nevoeiro é o quão persistente ele é. Para tanto, basta tomar a previsão para tantas horas quanto se queira como sendo a simples repetição da condição de tempo presente.

Para o caso desta pesquisa, onde a previsão do fenômeno nevoeiro está sendo abordada no formato categórico para até três horas, a previsão por persistência dá-se do seguinte modo: tomemos por x a variável hora, e por y a variável categórica visibilidade. Para visibilidades inferiores a 1000m, configura-se a formação de nevoeiro, e y recebe o valor 1; para visibilidades superiores ou iguais a 1000m, não há a condição de nevoeiro, e y recebe o valor -1.

Se em uma hora x qualquer existe a condição de nevoeiro, obtém-se o par $(x, 1)$. A previsão por persistência nos fornece a previsão para as horas $x+1$, $x+2$ e $x+3$ como sendo a manutenção da condição observada na hora x , gerando então os pares $(x+1, 1)$, $(x+2, 1)$ e $(x+3, 1)$. A inferência é realizada do mesmo modo para os casos em que não há a condição de nevoeiro, ou seja $y=-1$. A Tabela 18, abaixo, apresenta um exemplo da aplicação desta técnica de previsão.

TABELA 18- EXEMPLO DE PREVISÃO POR PERSISTÊNCIA

Hora	Previsões			
	0h	+1h	+2h	+3h
00	+1			
01	+1	+1		
02	+1	+1	+1	
03	+1	+1	+1	+1
04	-1	+1	+1	+1
05	-1	-1	+1	+1
06	-1	-1	-1	+1
07	-1	-1	-1	-1
08	-1	-1	-1	-1

FONTE: A Autora, 2016.

onde a coluna “0h” apresenta o fenômeno observado. Nota-se que o erro das previsões concentra-se nas horas em que ocorre a mudança da condição de nevoeiro para não-nevoeiro, ou vice-versa. Assim, quanto menos variações houver nas condições da série temporal, menos erros haverá nas inferências.

Pode-se dizer que o fenômeno nevoeiro apresenta comportamento bastante estável, no sentido de que uma vez instaurado, a condição de restrição de visibilidade se mantém por várias horas. Esta condição gera poucas variações na série temporal, o que pode proporcionar às previsões por persistência um bom desempenho para este fenômeno. Este experimento será realizado posteriormente, na Seção 4.2.

3.5 ESPECIFICAÇÕES DA REDE NEURAL UTILIZADA

Foi utilizada uma Rede Neural Artificial RBF, cuja função de base radial é do tipo gaussiana, e a função de ativação na camada de saída é do tipo tangente hiperbólica. Esta função foi escolhida para a última camada, pois é uma das funções contínuas que levam à saturação do neurônio de saída. O número de dados do vetor de entrada, bem como o número de neurônios ocultos variou nas diferentes etapas do treinamento, como será descrito na Seção 4.3.1. Na camada de saída, foi utilizado apenas um neurônio, visto que a previsão do fenômeno nevoeiro, nesta pesquisa, será realizada na forma categórica.

O programa foi executado em um computador com processador Intel i7, com memória RAM de 8Gb. O código do programa foi cedido pelo Professor Ricardo Almeida, e foi escrito em linguagem computacional FORTRAN 95.

Foram usados 75% dos dados da amostra para o treinamento da rede, 15% para os testes, e 10% na validação.

A partir do conjunto de treinamento são ajustados os centros e os desvios padrões de cada classe, definindo as RBF, bem como são ajustados os pesos das ligações sinápticas da camada de saída. Esses dados da amostra foram normalizados através dos valores máximos e mínimos encontrados: o valor máximo foi tomado como 1, e o valor mínimo como -1. Depois de normalizados os dados, foram utilizadas 10 iterações na validação cruzada. Ou seja, a rede separou a amostra completa em 10 partes, reorganizou os vetores de forma aleatória, e realizou um treinamento independente com cada uma das amostras menores.

O conjunto de testes é apresentado após cada época de treinamento para que se identifique a ocorrência de *overtraining*, caracterizada pelo aumento do erro das previsões com o conjunto de teste.

Ao final do treinamento, é apresentado à rede o conjunto de validação, para que seja avaliado o desempenho da configuração obtida. Como foram utilizadas 10 RNAs distintas, avaliou-se o desempenho de cada uma, e como desempenho final, apresentou-se as médias dos desempenhos dos treinamentos.

O número de épocas de treinamento, na primeira e na segunda fase do treinamento da rede, variaram de acordo com o número de variáveis de entrada e o número neurônios ocultos, mas sempre foram estabelecidas de modo que a máxima movimentação da posição dos centros das classes, durante o processo de ajuste, fosse inferior, ou igual, a 1×10^{-6} . O fator de escala dos pesos iniciais das sinapses foi considerado como 0,25 e as taxas de aprendizagem na correção das variâncias, na correção das posições dos centros e na correção dos pesos foi de 0,007.

3.6 SELEÇÃO DE DADOS DE ENTRADA PARA A RNA

Dentre todas as variáveis disponíveis na amostra de dados, era necessário escolher as mais relevantes para os experimentos com a RNA. O método adotado para a realização da escolha foi a análise da correlação linear entre os valores de cada uma das variáveis, e os valores de visibilidade. A variável visibilidade foi adotada, inicialmente, como parâmetro. Os experimentos mostraram que ela é a que determina de forma mais direta o fenômeno de interesse desta pesquisa.

Além disso, de todo o período de valores que compunha a amostra, era necessário determinar quão longas seriam as séries temporais utilizadas como valores de entrada para rede.

Assim, foi calculada a correlação entre cada uma das variáveis e a variável visibilidade, de 0 a 12 horas anteriores, através do Coeficiente de Pearson, indicado a seguir.

$$r = \frac{n \sum (v_i \cdot p_i) - (\sum v_i)(\sum p_i)}{\sqrt{n \sum v_i^2 - (\sum v_i)^2} \cdot \sqrt{n \sum p_i^2 - (\sum p_i)^2}} \quad (12)$$

onde r é o valor da correlação, o qual varia entre -1 e 1; n é o número de elementos da amostra; v_i representa o valor da visibilidade na hora; e p_i indica o valor de cada uma das demais variáveis. Para cada variável e hora anterior foi calculado um coeficiente, cujos resultados estão dispostos no Anexo A, e ilustrados na Figura 17, a seguir.

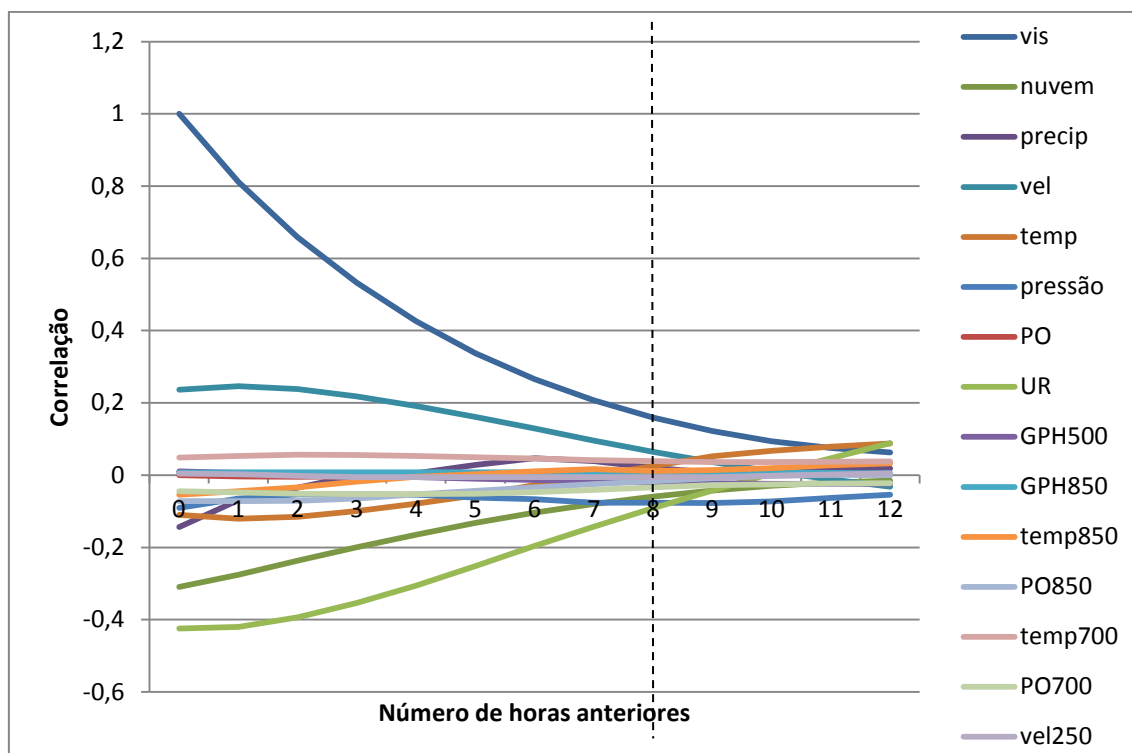


FIGURA 17- EVOLUÇÃO TEMPORAL DA CORRELAÇÃO ENTRE AS VARIÁVEIS COMPONENTES DA AMOSTRA E A VARIÁVEL VISIBILIDADE
 FONTE: A Autora, 2016.

É possível notar que, das variáveis selecionadas, a que possui maior correlação com a visibilidade é a própria visibilidade, e que os valores de correlação mais significativos se concentram no período de 1 a 8 horas anteriores. Por isso, atribuiu-se o horário de 8 horas como o horário de corte para as séries temporais.

Assim, os testes com a RNA iniciaram com séries temporais contendo de uma a oito horas de dados de visibilidade, e posteriormente foram introduzidas outras variáveis. Os resultados desses experimentos serão descritos na Seção 4.3.

4 EXPERIMENTOS DE PREVISÃO DE OCORRÊNCIA DE NEVOEIRO

4.1 PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DE CLIMATOLOGIA

Para realizar a previsão de um fenômeno através de climatologia, adotou-se a abordagem de buscar nos eventos passados características comuns a todas as ocorrências. A partir disso, inferiu-se que, na configuração destas determinadas características, ocorreria o fenômeno em questão.

No caso desta pesquisa, deseja-se prever a ocorrência, ou não, do fenômeno nevoeiro. O levantamento sobre as condições reinantes no momento de ocorrência do fenômeno foi realizado na Seção 3.4 e, a partir daquelas análises, chegou-se às informações apresentadas Tabela 17, que esquematiza as situações favoráveis à ocorrência do evento.

A previsão para uma, duas e três horas foi então realizada, inferindo-se que, na ocorrência simultânea de tais condições, haveria a formação de nevoeiro nas três horas subsequentes. Os resultados das previsões foram validados pelas métricas utilizadas na avaliação de desempenho descritas na Seção 3.3, e podem ser observadas na Tabela 19, a seguir.

TABELA 19- DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIRO POR CLIMATOLOGIA

Horas de previsão	+1h	PC	0,95	ISC	0,04	V	6,46	RFA	0,96	PD	0,28
	+2h	PC	0,95	ISC	0,03	V	6,46	RFA	0,96	PD	0,24
	+3h	PC	0,95	ISC	0,03	V	6,46	RFA	0,97	PD	0,21

FONTE: A Autora, 2016.

É possível perceber que a Proporção Correta é bastante alta. No entanto, a Razão de Falso Alarme também o é, chegando próxima ao valor 1. Estes resultados indicam que, utilizando esta técnica de previsão, tende-se a superestimar a ocorrência de nevoeiros. Esta constatação fica clara através do índice Viés, pois seu valor mostra que o evento é previsto, aproximadamente, seis vezes mais frequentemente do que de fato ocorre.

É possível restringir ainda mais os intervalos admitidos, buscando um maior refinamento da previsão. No entanto, é importante salientar que a climatologia verifica as condições reinantes quando na ocorrência do

fenômeno, mas a configuração destas condições não, necessariamente, implica na ocorrência do fenômeno.

4.2 PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DE PERSISTÊNCIA

Na Seção 3.5 foi realizada uma descrição a respeito de variáveis meteorológicas persistentes, e de como se dá a técnica de previsão por persistência.

No caso desta pesquisa, onde a previsão do fenômeno nevoeiro dar-se-á apenas no formato categórico, a previsão por persistência ocorreu nos moldes do exemplo disposto na Tabela 18.

Foram utilizados os dados de visibilidade contidos nas mensagens METAR, no período de 00h 00min do dia 01/01/2015 à 23h 00min do dia 31/12/2015, disponíveis na REDEMET (www.redemet.aer.mil.br). Estes dados foram utilizados para a aplicação desta técnica de previsão, pois também o foram na operação final da RNA, possibilitando, assim, uma melhor comparação entre as técnicas. Da mesma forma que no tratamento dos dados para o treinamento, as informações especiais e locais foram descartadas, considerando-se apenas as informações regulares. No código meteorológico METAR, visibilidades superiores ou iguais a 10km são codificadas como “9999”, e as visibilidades ausentes, codificadas como “////”. Para não haver confusão com a notação adotada anteriormente, estas foram substituídas por 10km, e “9999”, respectivamente. Para os casos de codificação “CAVOK”², a visibilidade foi adotada também, como 10km.

Todas as informações de visibilidade foram tomadas no formato bipolar, e a previsão para uma, duas e três horas consistiu, apenas, em repetir a condição presente. Comparando os resultados das previsões aos valores de visibilidade da amostra, tal como no esquema apresentado na Tabela 18, foi possível verificar a acurácia deste método de previsão. As métricas utilizadas, para tanto, foram as mesmas apresentadas na Seção 3.3, e podem ser observadas na Tabela 20, a seguir.

² A sigla “CAVOK” substitui, no código METAR, as informações de visibilidade, tempo presente, e cobertura de nuvens, quando estas, simultaneamente, não têm significado operacional.

TABELA 20- DESEMPENHO DA TÉCNICA DE PREVISÃO DE NEVOEIROS POR PERSISTÊNCIA

Horas de previsão	+1h	PC	0,99	ISC	0,61	V	1	RFA	0,24	PD	0,76
	+2h	PC	0,98	ISC	0,42	V	1	RFA	0,41	PD	0,59
	+3h	PC	0,97	ISC	0,32	V	1	RFA	0,52	PD	0,48

FONTE: A Autora, 2016.

Nota-se que Proporção Correta é bastante elevada. Este resultado já era esperado, pois como já citado anteriormente, existe uma predominância das condições de não-nevoeiro em relação à ocorrência do evento. Observa-se também que o índice PD é bastante razoável, visto que esta é uma técnica de previsão sem destreza. A RFA, no entanto, é bastante elevada já para a primeira hora de previsão, e aumenta ainda mais na segunda e terceira horas.

Um índice interessante nesta técnica de previsão é o Viés, pois, para todas as previsões, obteve-se valor 1. Este valor indica que o fenômeno foi previsto exatamente o mesmo número de vezes que ocorreu. Isso ocorre porque a previsão através do método de persistência consiste em, apenas, projetar a condição de interesse, sem alterar seu valor, mantendo o número de previsões igual ao de ocorrências.

4.3 PREVISÃO DO FENÔMENO NEVOEIRO ATRAVÉS DA RNA RBF

Para a realização das previsões para três horas subsequente foram utilizadas três Redes Neurais distintas: uma para a previsão para uma hora (R1), uma para a previsão para duas horas (R2), e uma para a previsão para três horas (R3). Cada uma delas foi treinada e analisada de maneira independente das demais. O processo de treinamento das RNAs passou por várias etapas até se chegar nas configurações ótimas, conforme descrito a seguir.

4.3.1 Treinamento das Redes Neurais Artificiais

Como já mencionado, os primeiros experimentos foram realizados apenas com informações de visibilidade como dados de entrada. Realizaram-se oito treinamentos para cada RNA, cada treinamento com uma série temporal

distinta, com número de dados de visibilidade em horários anteriores à ocorrência do nevoeiro variando entre um e oito.

Tomando o melhor resultado dentre as séries temporais de visibilidade, foi realizado o ajuste do número de neurônios ocultos de cada uma das redes. Encontrado o melhor resultado desta etapa, avaliou-se a combinação da melhor configuração obtida na fase anterior com as demais variáveis, uma de cada vez. Tomando o melhor resultado, novamente foi realizado o ajuste de neurônios ocultos para cada Rede. Para a melhor configuração obtida, acrescentou-se mais uma variável, e assim sucessivamente, até que a Rede não apresentasse melhora nos desempenhos. Obtida a melhor combinação entre as variáveis, foi realizado o ajuste da amostra dos dados de treinamento, e, obtida a melhor proporção, realizado novamente o ajuste do número de neurônios ocultos.

Ao final de cada uma das etapas do treinamento das Redes Neurais, será exibida uma tabela, apresentando a melhor configuração das RNA obtidas até aquele momento. A melhor configuração em cada etapa é a escolhida para prosseguir para o próximo ajuste. Todo o processo será descrito, detalhadamente, a seguir.

4.3.1.1 Primeira etapa do treinamento: Treinamento com as séries temporais de visibilidade

Os experimentos se iniciaram com o treinamento empregando registros constituídos apenas de dados de visibilidade, nos formatos bipolar e métrico. Esses registros continham entre uma e oito informações de visibilidade, as quais são apresentadas à RNA no formato de um vetor n -dimensional. A camada intermediária desta RNA inicial dispunha de cinco neurônios ocultos.

Para os testes realizados com os dados de visibilidade no formato categórico foram obtidos os resultados apresentados na Tabela 21.

TABELA 21- RESULTADOS DOS EXPERIMENTOS COM SÉRIES TEMPORAIS CONSTITUÍDAS DE DADOS DE VISIBILIDADE NO FORMATO BIPOLAR

		vis							
Horas anteriores		1	2	3	4	5	6	7	8
R1	PC	0,16	0,16	0,91	0,91	0,91	0,91	0,91	0,91
	ISC	0,16	0,16	0,55	0,56	0,55	0,55	0,55	0,55
	V	6,46	6,46	0,96	0,96	0,96	0,96	0,96	0,96
	RFA	0,85	0,85	0,27	0,27	0,27	0,27	0,27	0,27
	PD	1	1	0,7	0,71	0,7	0,7	0,7	0,7
R2	PC	0,16	0,16	0,87	0,87	0,87	0,87	0,87	0,87
	ISC	0,16	0,16	0,4	0,4	0,39	0,39	0,39	0,39
	V	6,45	6,45	0,89	0,89	0,88	0,88	0,88	0,84
	RFA	0,85	0,85	0,4	0,38	0,4	0,4	0,4	0,39
	PD	1	1	0,54	0,54	0,53	0,53	0,53	0,51
R3	PC	0,16	0,16	0,68	0,85	0,85	0,86	0,86	0,86
	ISC	0,16	0,16	0,19	0,3	0,22	0,25	0,26	0,26
	V	6,46	6,45	0,47	0,7	0,47	0,56	0,6	0,59
	RFA	0,85	0,85	0,38	0,46	0,45	0,44	0,45	0,44
	PD	1	1	0,25	0,4	0,26	0,31	0,33	0,33

FONTE: A Autora, 2016.

A convergência no treinamento das RNAs pode ser observada nas Figuras 18, 19 e 20, a seguir, as quais mostram a evolução dos erros dos conjuntos de treinamento em função do número de iterações.

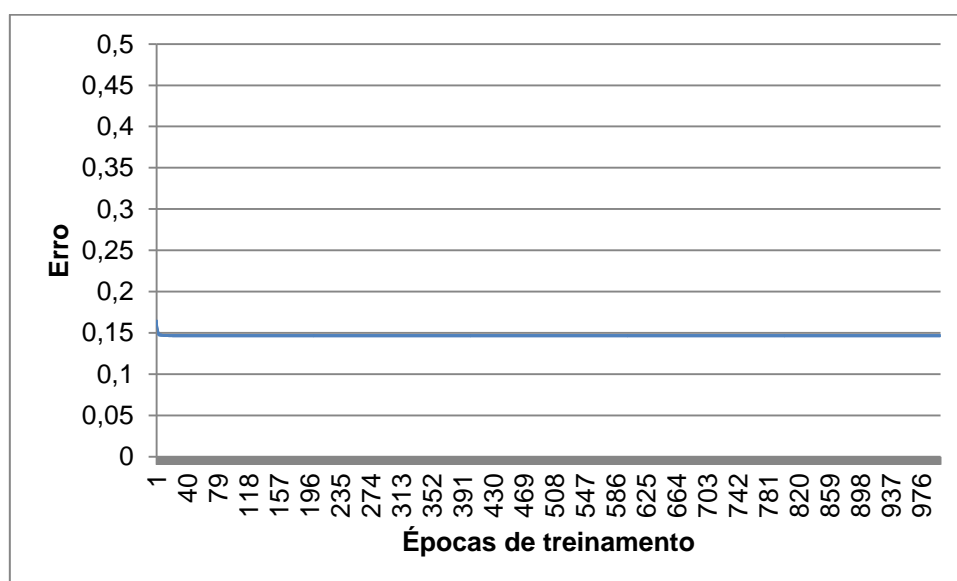


FIGURA 18- CONVERGÊNCIA DO ERRO DA REDE R1 NA PRIMEIRA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

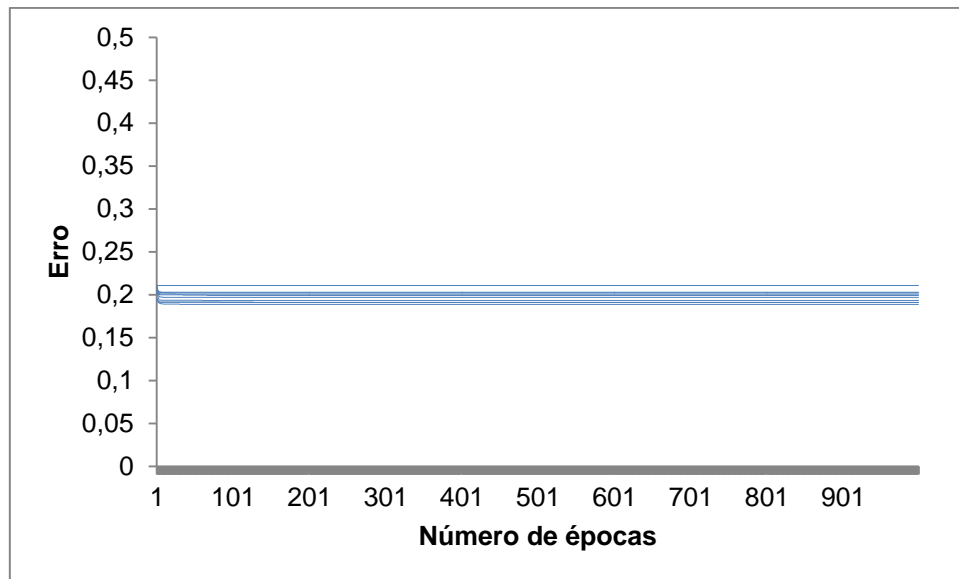


FIGURA 19- CONVERGÊNCIA DO ERRO DA REDE R2 NA PRIMEIRA FASE DE TREINAMENTO
 FONTE: A Autora, 2016.

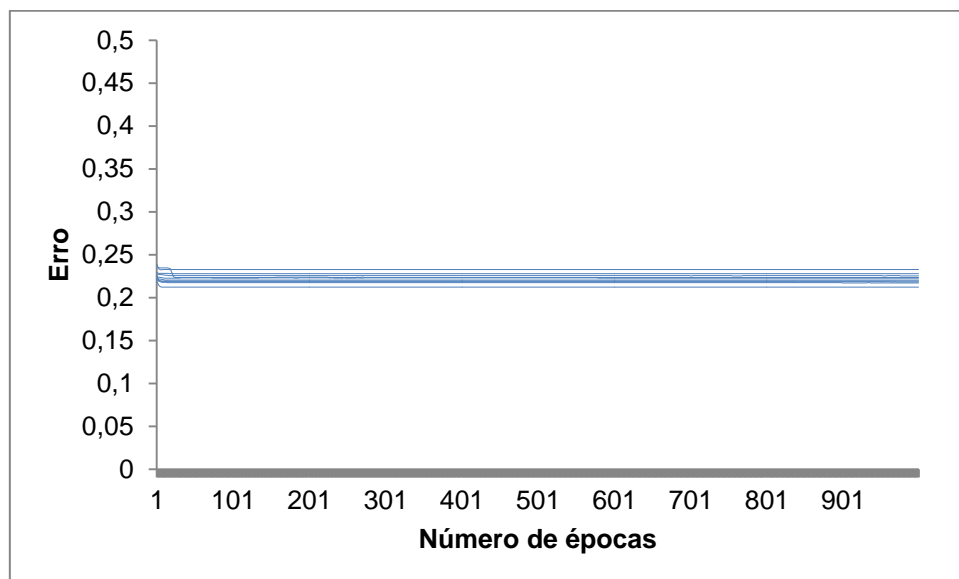


FIGURA 20- CONVERGÊNCIA DO ERRO DA REDE R3 NA PRIMEIRA FASE DE TREINAMENTO
 FONTE: A Autora, 2016.

Note-se que após uma rápida variação nas primeiras iterações, os erros tendem a se estabilizar, indicando que a rede atingiu uma posição de mínimo na superfície de erro e ali se manteve. Deve-se observar também que cada linha do gráfico representa uma determinada rede com um conjunto de dados distinto das demais e que, de forma geral, todas convergem para o mesmo mínimo, sugerindo que houve uma adequada convergência no

treinamento. Esse comportamento foi observado em todos os experimentos com as RNAs RBF utilizadas neste estudo.

Seguindo-se o critério de hierarquia entre os índices de desempenho descrito na Seção 3.3, observa-se que os melhores resultados obtidos para R1, R2 e R3 foram séries temporais com quatro valores de visibilidade, de quatro horas anteriores, no formato bipolar. Conforme pode ser observado na Tabela 22, abaixo, a qual descreve a melhor configuração obtida nesta etapa do treinamento.

TABELA 22- MELHOR DESEMPENHO DA PRIMEIRA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis
R1	4	vis
R2	4	vis
R3	4	vis

FONTE: A Autora, 2016.

Para os testes realizados com os dados de visibilidade no formato métrico, as RNAs não foram capazes de prever nenhum fenômeno de nevoeiro corretamente para nenhuma das três horas subsequentes. Este fato pode indicar uma dificuldade da RNA em agrupar corretamente os vetores, quando estes lhe são apresentados com uma gama maior de valores.

Obtida então a série temporal de dados de visibilidade que fornecem o melhor desempenho para R1, R2 e R3, passou-se ao ajuste do número de neurônios ocultos.

4.3.1.2 Segunda etapa do treinamento: Primeiro ajuste do número de neurônios ocultos

A partir do melhor desempenho obtido na etapa anterior, passou-se ao ajuste do número de neurônios ocultos. Nesta etapa, cada uma das Redes R1, R2 e R3 foi novamente treinada, com as mesmas informações de entrada, mas com variações no número de neurônios ocultos. Obtiveram-se os resultados dispostos na Tabela 23, a seguir, considerando-se o número de neurônios ocultos variando entre 2 e 10.

TABELA 23- PRIMEIRO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS

Número de neurônios ocultos		2	3	4	5	6	7	8	9	10
R1	PC	0,91	0,89	0,91	0,91	0,91	0,91	0,91	0,91	0,91
	ISC	0,55	0,51	0,54	0,56	0,55	0,55	0,55	0,55	0,55
	V	0,96	1,15	1	0,96	0,96	0,96	0,96	0,96	0,96
	RFA	0,27	0,35	0,29	0,27	0,27	0,27	0,27	0,27	0,27
	PD	0,7	0,72	0,7	0,71	0,7	0,7	0,7	0,7	0,7
R2	PC	0,87	0,86	0,87	0,87	0,87	0,87	0,87	0,87	0,87
	ISC	0,4	0,38	0,4	0,4	0,4	0,4	0,4	0,4	0,4
	V	0,9	1,04	0,93	0,89	0,9	0,89	0,9	0,89	0,9
	RFA	0,4	0,45	0,41	0,38	0,4	0,4	0,4	0,4	0,4
	PD	0,5	0,56	0,54	0,54	0,54	0,54	0,54	0,54	0,54
R3	PC	0,26	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85
	ISC	0,08	0,26	0,25	0,3	0,27	0,27	0,26	0,27	0,26
	V	0,19	0,72	0,63	0,7	0,68	0,67	0,65	0,67	0,62
	RFA	0,14	0,49	0,47	0,46	0,47	0,48	0,47	0,47	0,47
	PD	0,10	0,36	0,33	0,4	0,36	0,35	0,34	0,35	0,33

FONTE: A Autora, 2016.

Considerando o desempenho das variáveis PD e RFA, é possível perceber que os melhores desempenhos das Redes R1, R2 e R3, foram obtidos com a manutenção de cinco neurônios ocultos em todas as RNAs pois, muito embora R1 e R2 apresentassem valores de PD maiores com 3 neurônios ocultos, a RFA mostrava-se muito superior. Assim, para a utilização de três neurônios ocultos em R1 e R2, um pequeno acréscimo em PD acarretaria em um grande aumento em RFA, o que, como já mencionado, não é interessante em previsões meteorológicas.

A convergência no treinamento das RNAs pode ser observada nas Figuras 18, 19 e 20, já apresentadas, visto que a configuração nesta etapa do treinamento se manteve igual à configuração da etapa anterior, a saber, 5 neurônios na camada oculta. .Abaixo, a Tabela 24 descreve a melhor configuração obtida nesta etapa do treinamento.

TABELA 24- MELHOR DESEMPENHO DA SEGUNDA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis	Neurônios ocultos
R1	4	vis	5
R2	4	vis	5
R3	4	vis	5

FONTE: A Autora, 2016.

Definida a nova configuração das RNAs, passou-se então à próxima etapa do treinamento, em que as demais variáveis serão combinadas ao melhor desempenho até aqui obtido.

4.3.1.3 Terceira etapa do treinamento: Primeira inserção de novas variáveis

Realizaram-se os experimentos, combinando, uma a uma, todas as variáveis pertencentes à amostra à variável visibilidade. A inserção de novas variáveis foi realizada, arbitrando-se que as séries temporais de cada uma delas teria o mesmo tamanho da série temporal de visibilidades já definida. Os resultados destes experimentos podem ser observados na Tabela 25, a seguir:

TABELA 25- PRIMEIRA INSERÇÃO DE VARIÁVEIS

		vis+pressão	vis+precip	vis+vel	vis+temp	vis+PO	vis+UR	vis+nuvem	vis+vel250	vis+GPH500	vis+GPH850	vis+UR700	vis+UR500	vis+dir
R1	PC	0,91	0,91	0,91	0,91	0,91	0,91	0,92	0,91	0,91	0,91	0,91	0,91	0,91
	ISC	0,54	0,56	0,55	0,55	0,55	0,55	0,54	0,55	0,55	0,55	0,55	0,55	0,56
	V	0,9	0,96	0,93	0,96	0,96	0,93	0,95	0,95	0,95	0,96	0,93	0,96	0,94
	RFA	0,25	0,27	0,26	0,28	0,28	0,27	0,28	0,27	0,27	0,27	0,27	0,27	0,27
	PD	0,67	0,7	0,68	0,69	0,69	0,68	0,69	0,69	0,7	0,7	0,68	0,7	0,7
R2	PC	0,88	0,87	0,88	0,87	0,87	0,88	0,88	0,87	0,87	0,87	0,87	0,87	0,87
	ISC	0,37	0,4	0,39	0,38	0,38	0,39	0,37	0,38	0,39	0,4	0,4	0,4	0,4
	V	0,66	0,89	0,83	0,83	0,85	0,85	0,76	0,85	0,86	0,86	0,89	0,89	0,8
	RFA	0,31	0,38	0,39	0,39	0,4	0,39	0,37	0,4	0,4	0,54	0,4	0,4	0,37
	PD	0,45	0,54	0,51	0,51	0,51	0,52	0,48	0,51	0,52	0,4	0,54	0,54	0,5

		vis+pressão	vis+precip	vis+vel	vis+temp	vis+PO	vis+UR	vis+nuvem	vis+vel250	vis+GPH500	vis+GPH850	vis+UR700	vis+UR500	vis+dir
R3	PC	0,87	0,85	0,86	0,85	0,85	0,85	0,87	0,85	0,85	0,85	0,51	0,51	0,85
	ISC	0,26	0,3	0,26	0,23	0,23	0,24	0,25	0,25	0,26	0,26	0,15	0,14	0,29
	V	0,46	0,75	0,57	0,52	0,53	0,57	0,54	0,6	0,58	0,57	0,38	0,34	0,64
	RFA	0,33	0,46	0,43	0,45	0,46	0,45	0,43	0,47	0,44	0,44	0,28	0,28	0,43
	PD	0,3	0,4	0,33	0,28	0,29	0,31	0,3	0,32	0,32	0,32	0,2	0,18	0,36

FONTE: A Autora, 2016.

O treinamento das RNAs R1, R2 e R3 foi realizado e, foi obtida a convergência do treinamento expressa pela estabilização dos erros do treinamento, conforme pode ser observado nas Figuras 21, 22 e 23, a seguir.

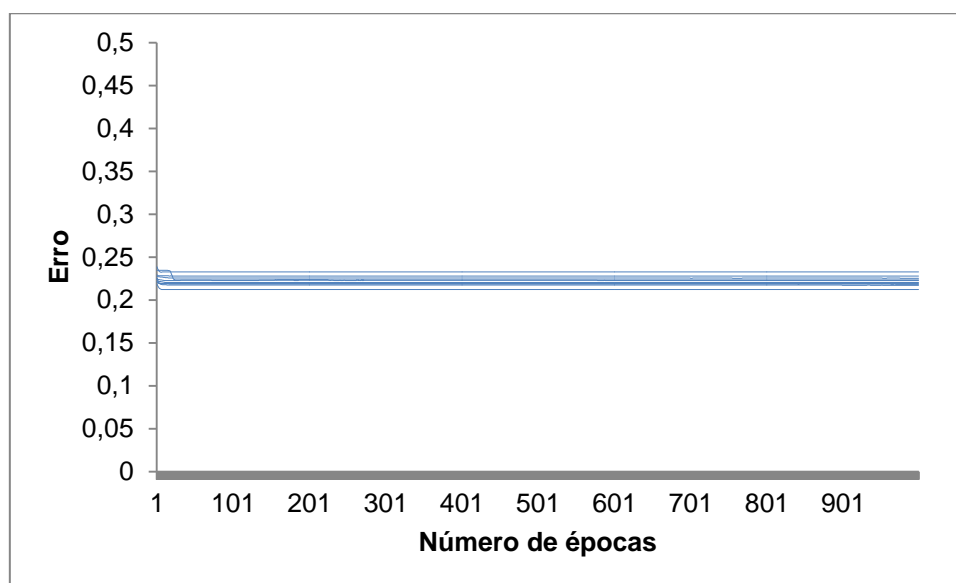


FIGURA 21- CONVERGÊNCIA DO ERRO DA REDE R1 NA TERCEIRA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

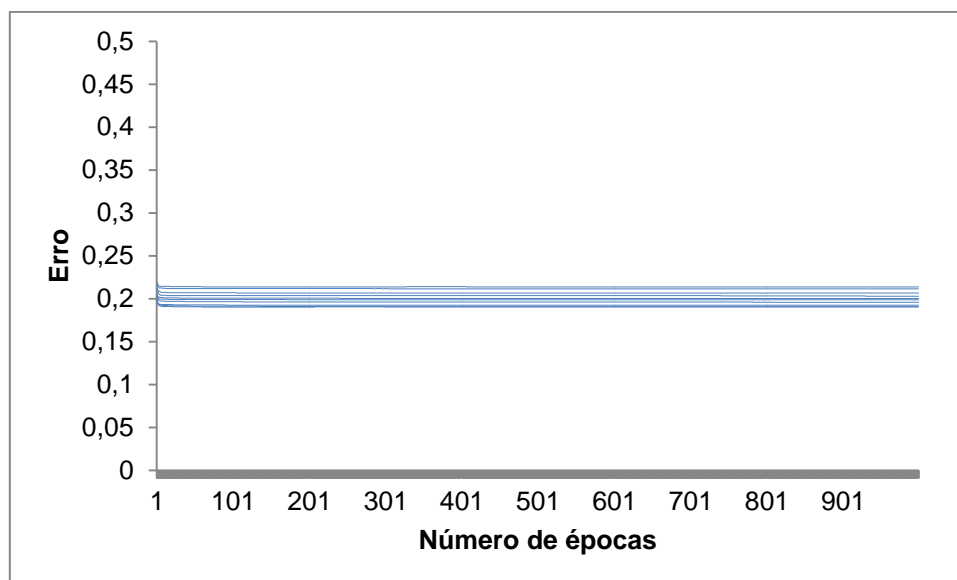


FIGURA 22- CONVERGÊNCIA DO ERRO DA REDE R2 NA TERCEIRA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

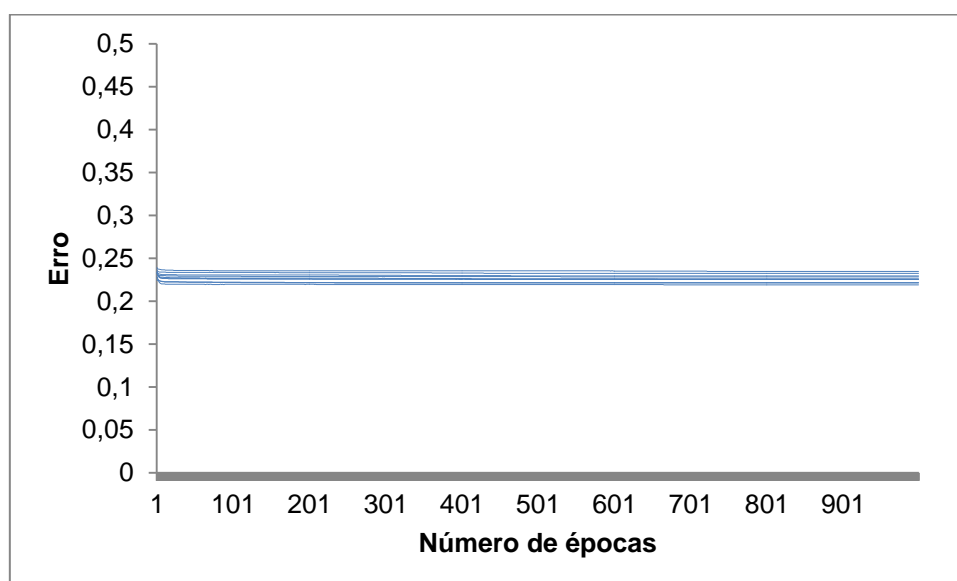


FIGURA 23- CONVERGÊNCIA DO ERRO DA REDE R3 NA TERCEIRA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

Observa-se que, para as três RNAs, a variável que apresentou melhor combinação com a variável visibilidade foi a variável precipitação. Abaixo, a Tabela 26 descreve a melhor configuração obtida nesta etapa do treinamento.

TABELA 26- MELHOR DESEMPENHO DA TERCEIRA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis	Neurônios ocultos
R1	4	Vis+precip	5
R2	4	Vis+precip	5
R3	4	Vis+precip	5

FONTE: A Autora, 2016.

Nota-se que, apesar de desta combinação ter gerado resultados superiores às demais combinações, o desempenho das RNAs, em comparação à etapa de treinamento anterior não apresentou melhora. Uma razão plausível para este resultado é que, com a inserção de mais informações de entrada, a rede necessite de mais neurônios para processá-los. Passou-se, então, ao segundo ajuste no número de neurônios ocultos.

4.3.1.4 Quarta etapa do treinamento: Segundo ajuste do número de neurônios ocultos

Realizou-se um novo ajuste do número de neurônios ocultos, a fim de verificar-se a hipótese levantada ao final da etapa de treinamento anterior. Seguiu-se a mesma estratégia adotada no primeiro ajuste, em que, as informações de entrada permanecem as mesmas, e varia-se apenas o número de neurônios ocultos. Supondo que, com a inserção de mais informações, a rede necessite de mais neurônios para processamento, o número de neurônios ocultos foi variado, então, a partir do valor que já havia sido determinado anteriormente. Assim, os experimentos foram realizados com a variação entre 5 e 10 neurônios inseridos na camada oculta.

Os resultados obtidos podem ser observados na Tabela 27.

TABELA 27- SEGUNDO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS

Número de neurônios ocultos		5	6	7	8	9	10
R1	PC	0,91	0,91	0,91	0,91	0,91	0,91
	ISC	0,56	0,56	0,56	0,56	0,56	0,56
	V	0,96	0,96	0,96	0,96	0,96	0,96
	RFA	0,27	0,27	0,27	0,27	0,27	0,27
	PD	0,7	0,7	0,7	0,7	0,7	0,7

Número de neurônios ocultos		5	6	7	8	9	10
R2	PC	0,87	0,87	0,87	0,87	0,87	0,87
	ISC	0,4	0,4	0,4	0,4	0,4	0,4
	V	0,89	0,89	0,89	0,89	0,89	0,89
	RFA	0,38	0,38	0,38	0,38	0,38	0,38
	PD	0,54	0,54	0,54	0,54	0,54	0,54
R3	PC	0,85	0,85	0,85	0,85	0,85	0,85
	ISC	0,3	0,3	0,3	0,3	0,3	0,3
	V	0,75	0,75	0,75	0,75	0,75	0,75
	RFA	0,46	0,46	0,46	0,46	0,46	0,46
	PD	0,4	0,4	0,4	0,4	0,4	0,4

FONTE: A Autora, 2016.

É possível observar que a inserção de mais neurônios ocultos não trouxe melhora no desempenho das RNAs. Nota-se que os índices permaneceram os mesmos, o que indica que cinco neurônios ocultos são ótimos para a esta etapa do treinamento. Supõe-se que os índices não se alteraram porque nenhum vetor foi associado aos novos centros, ou porque, durante o processo de ajuste da posição dos centros das classes, os centros das classes acrescentadas foram sendo aproximados de outros centros já existentes, de modo que acabaram se sobrepondo. Desta forma, apesar de inicialmente haver mais de cinco neurônios ocultos, com o ajuste das posições dos centros das classes, a camada oculta efetivamente dispôs de apenas cinco neurônios. Assim, manteve-se a configuração anterior, com cinco neurônios ocultos, conforme pode ser observado na Tabela 28.

TABELA 28- MELHOR DESEMPENHO NO SEGUNDO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS

RNA	Horas anteriores	Variáveis	Neurônios ocultos
R1	4	Vis+precip	5
R2	4	Vis+precip	5
R3	4	Vis+precip	5

FONTE: A Autora, 2016.

A convergência dos erros de treinamento das Redes R1, R2 e R3, nesta fase de treinamento, podem ser observadas nas Figuras 21, 22, e 23,

anteriormente apresentadas, visto que se manteve a mesma configuração da etapa anterior para as três Redes Neurais.

Observa-se que mesmo com a inserção de novas variáveis e ajuste no número de neurônios ocultos, as Redes Neurais Artificiais não apresentaram melhora de desempenho. A partir disso, supõe-se que, a inserção de mais variáveis, e novos ajustes na camada oculta também não trarão resultados significativos. Encerraram-se, então, os experimentos com inserção de novas variáveis³.

Como é possível observar na Tabela 25, a combinação da variável precipitação, com a variável visibilidade, apresentou-se melhor do que as demais combinações. No entanto, não houve melhora no desempenho da rede em comparação aos experimentos realizados apenas com informações de visibilidades fornecidas às RNAs, conforme pode ser observado na Tabela 21. Observa-se, inclusive, uma pequena piora no desempenho das RNAs quando na inserção da nova variável. Por esta razão, dado o melhor desempenho observado, foram selecionadas, como entradas que produzem o melhor desempenho das RNAs, apenas as informações referentes à variável visibilidade.

Abaixo, a Tabela 29 descreve a melhor configuração obtida ao final dos experimentos com a inserção de novas variáveis.

TABELA 29- MELHOR DESEMPENHO DA QUARTA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis	Neurônios ocultos
R1	4	Vis	5
R2	4	Vis	5
R3	4	Vis	5

FONTE: A Autora, 2016.

Passou-se, então, à próxima etapa de treinamento, onde será variada a proporção entre os dados de treinamento da amostra.

³ Foram testadas outras combinações, com um número maior de variáveis. No entanto, as RNAs não apresentaram melhora em seus desempenhos, tal como na combinação apresentada.

4.3.1.5 Quinta etapa do treinamento: Ajuste da amostra de dados de treinamento

Por se tratar de um evento raro, o número de ocorrências do fenômeno nevoeiro é bastante inferior ao número de ocorrências de não-nevoeiro. Esta informação pode ser verificada na Tabela 30, a seguir, onde é apresentada a distribuição dos valores de visibilidade disponíveis na amostra utilizada nesta pesquisa.

TABELA 30- DISTRIBUIÇÃO DOS VALORES DE VISIBILIDADE

Intervalos de visibilidade (em metros)	Número de casos	Intervalos de visibilidade (em metros)	Número de casos
[0,1000)	7300	[10000,11000)	5447
[1000,2000)	1511	[11000,12000)	31
[2000,3000)	2458	[12000,13000)	6345
[3000,4000)	4253	[13000,14000)	308
[4000,5000)	3849	[14000,15000)	332
[5000,6000)	5586	[15000,16000)	9561
[6000,7000)	3934	[16000,17000)	177
[7000,8000)	2879	[17000,18000)	218
[8000,9000)	7078	[18000,19000)	636
[9000,10000)	2774	[19000,20000]	93006

FONTE: A Autora, 2016.

O fato de haver um número muito maior de casos de não-nevoeiro (aproximadamente vinte vezes mais) poderia estar dificultando a identificação do fenômeno de interesse pela Rede Neural. Para diminuir o número de casos de não-nevoeiro sem prejudicar a representatividade da amostra, foram tomados grupos de dados n vezes menores do que a amostra original, com n variando entre 2 e 20. Ajustes de razão maiores que 20 não foram adotados, pois o número de casos tornar-se-ia maior que o número de não-casos, fazendo-nos retornar ao problema original de desequilíbrio entre os grupos.

O ajuste foi realizado tomando-se os vetores com cinco informações de visibilidade, onde as quatro primeiras são as componentes do vetor de entrada que é apresentado à rede, e o quinto valor é o dado de referência fornecido como parâmetro de verificação da assertividade da previsão fornecida. Para o ajuste, foi observado apenas o quinto elemento do vetor supracitado.

		Razão de Ajuste									
		1	2	3	4	5	6	7	8	9	10
R3	PC	0,85	0,92	0,9	0,88	0,86	0,85	0,83	0,82	0,8	0,8
	ISC	0,28	0,33	0,35	0,37	0,38	0,39	0,4	0,41	0,41	0,42
	V	0,7	0,69	0,62	0,57	0,56	0,52	0,52	0,53	0,53	0,53
	RFA	0,47	0,4	0,32	0,25	0,23	0,19	0,17	0,15	0,16	0,14
	PD	0,37	0,42	0,42	0,43	0,43	0,43	0,43	0,46	0,45	0,45
		11	12	13	14	15	16	17	18	19	20
	PC	0,78	0,77	0,76	0,75	0,75	0,74	0,73	0,72	0,72	0,71
	ISC	0,42	0,43	0,43	0,43	0,43	0,44	0,43	0,44	0,44	0,44
	V	0,52	0,54	0,53	0,52	0,51	0,51	0,51	0,51	0,51	0,49
	RFA	0,14	0,14	0,14	0,12	0,11	0,1	0,1	0,1	0,09	0,08
	PD	0,45	0,46	0,46	0,46	0,46	0,46	0,46	0,46	0,46	0,46

FONTE: A Autora, 2016.

Analisando o desempenho de R1, R2 e R3 segundo a hierarquia entre métricas já estabelecida, pode-se notar que o melhor desempenho das RNAs foram para ajustes de razão 20. A convergência do erro dos conjuntos de treinamento nesta fase das simulações pode ser observado nas Figuras 24, 25 e 26, a seguir.

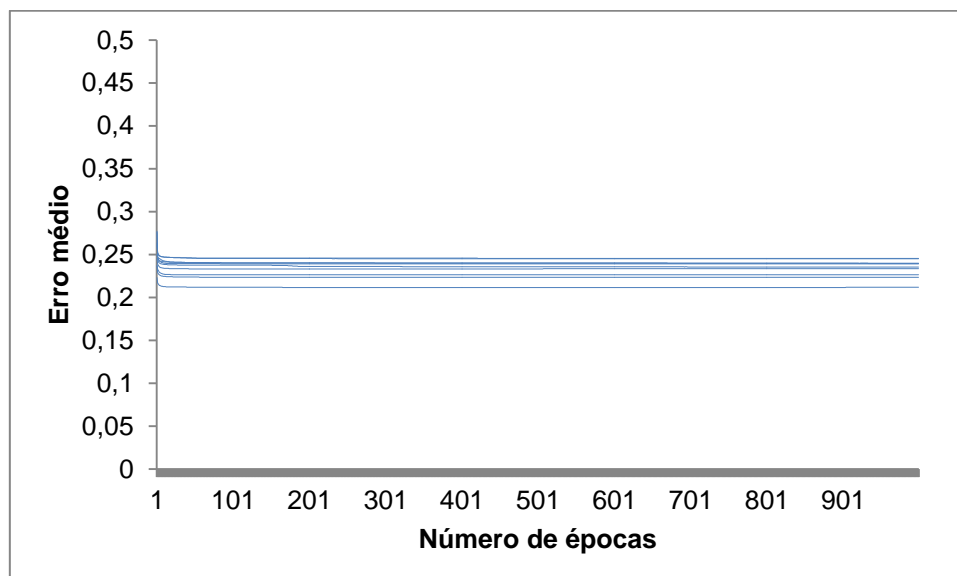


FIGURA 24- CONVERGÊNCIA DO ERRO DA REDE R1 NA QUINTA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

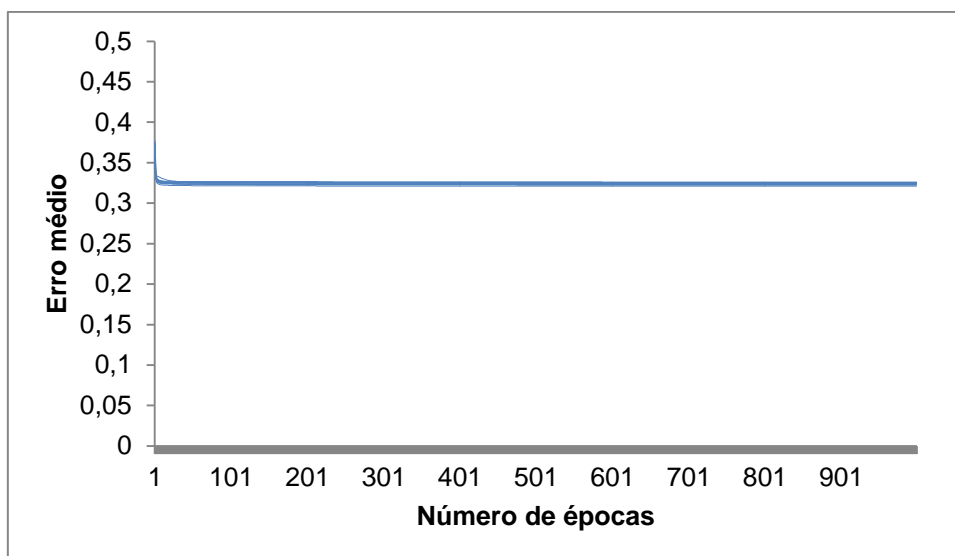


FIGURA 25- CONVERGÊNCIA DO ERRO DA REDE R2 NA QUINTA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

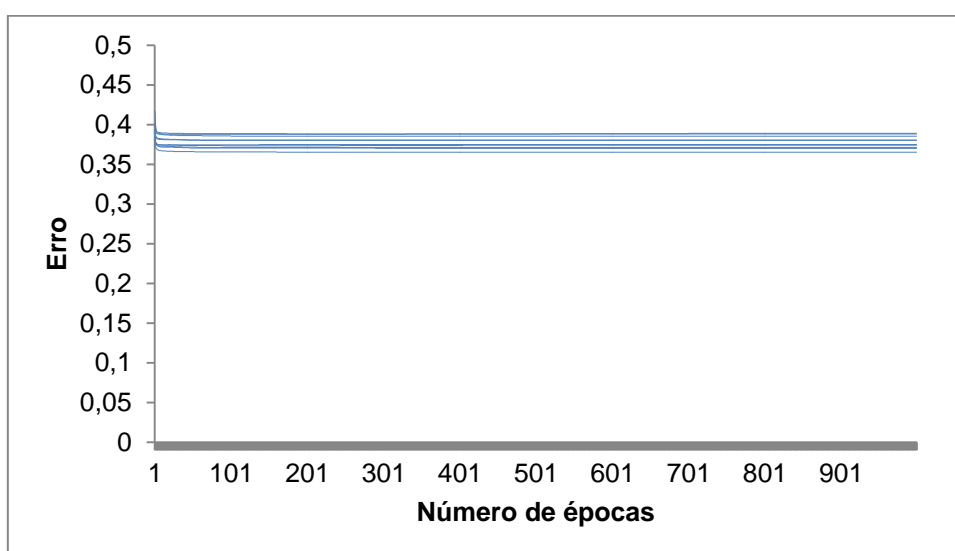


FIGURA 26- CONVERGÊNCIA DO ERRO DA REDE R3 NA QUINTA FASE DE TREINAMENTO

FONTE: A Autora, 2016.

A configuração da rede ao final desta etapa do treinamento pode ser observada na Tabela 32, abaixo.

TABELA 32- MELHOR DESEMPENHO DA QUINTA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis	Neurônios ocultos	Razão de Ajuste
R1	4	Vis	5	20
R2	4	Vis	5	20
R3	4	Vis	5	20

FONTE: A Autora, 2016.

Com a melhor razão de ajuste encontrada, passou-se à última etapa do treinamento das RNAs: o terceiro ajuste no número de neurônios ocultos.

4.3.1.6 Sexta etapa do treinamento: Terceiro ajuste do número de neurônios ocultos

Como última etapa do treinamento, foi realizado mais um ajuste no número de neurônios ocultos de cada uma das RNA. Os experimentos foram realizados, novamente, com a manutenção dos dados de entrada, e variação apenas do número de neurônios ocultos. A camada oculta recebeu de 2 a 5 neurônios ocultos, pois se partiu da premissa de que quanto maior o número de vetores a serem apresentados, maior o número de padrões de referência, os quais são representados pelos representantes centrais dos *clusters*; e, com a redução no número de dados na amostra de treinamento, é possível que um número inferior de neurônios seja necessário. Os resultados dos experimentos estão apresentados na Tabela 33.

TABELA 33- TERCEIRO AJUSTE DO NÚMERO DE NEURÔNIOS OCULTOS

Número de neurônios ocultos		2	3	4	5
R1	PC	0,85	0,85	0,85	0,85
	ISC	0,71	0,71	0,71	0,71
	V	0,77	0,78	0,77	0,76
	RFA	0,05	0,06	0,06	0,04
	PD	0,74	0,74	0,73	0,73
R2	PC	0,77	0,77	0,77	0,77
	ISC	0,56	0,56	0,56	0,56
	V	0,63	0,64	0,62	0,62
	RFA	0,07	0,08	0,06	0,07
	PD	0,59	0,59	0,58	0,58
R3	PC	0,71	0,71	0,71	0,71
	ISC	0,45	0,44	0,44	0,44
	V	0,52	0,52	0,51	0,49
	RFA	0,09	0,1	0,09	0,08
	PD	0,47	0,47	0,46	0,46

FONTE: A Autora, 2016.

Observa-se que os melhores desempenhos das três RNAs, R1, R2 e R3, após o ajuste dos dados da amostra de treinamento, ocorrem quando há apenas dois neurônios na camada oculta.

A convergência do erro para as Redes R1, R2 e R3 nesta fase do treinamento pode ser observada nas Figuras 27, 28 e 29, a seguir.

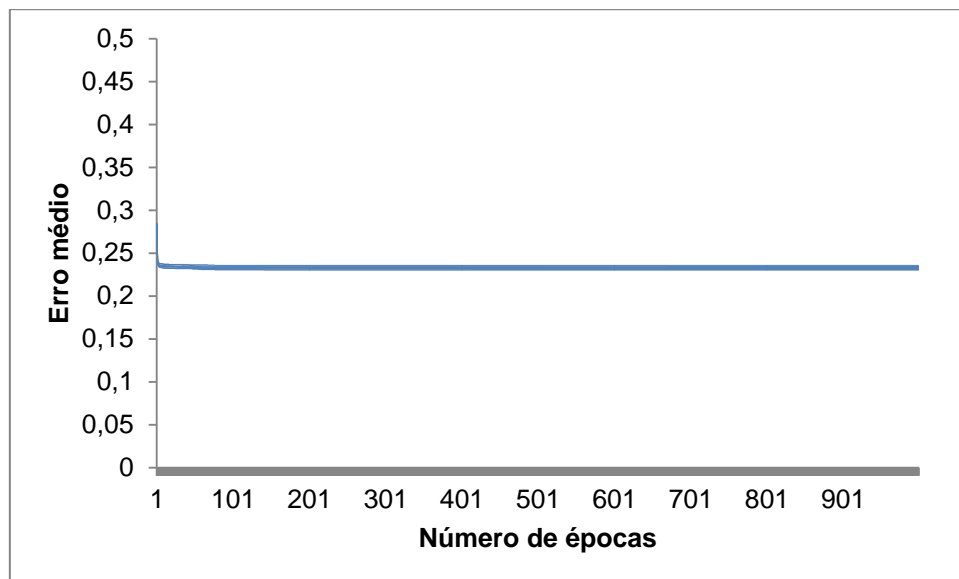


FIGURA 27- CONVERGÊNCIA DO ERRO DA REDE R1 NA SEXTA FASE DE TREINAMENTO
FONTE: A Autora, 2016.

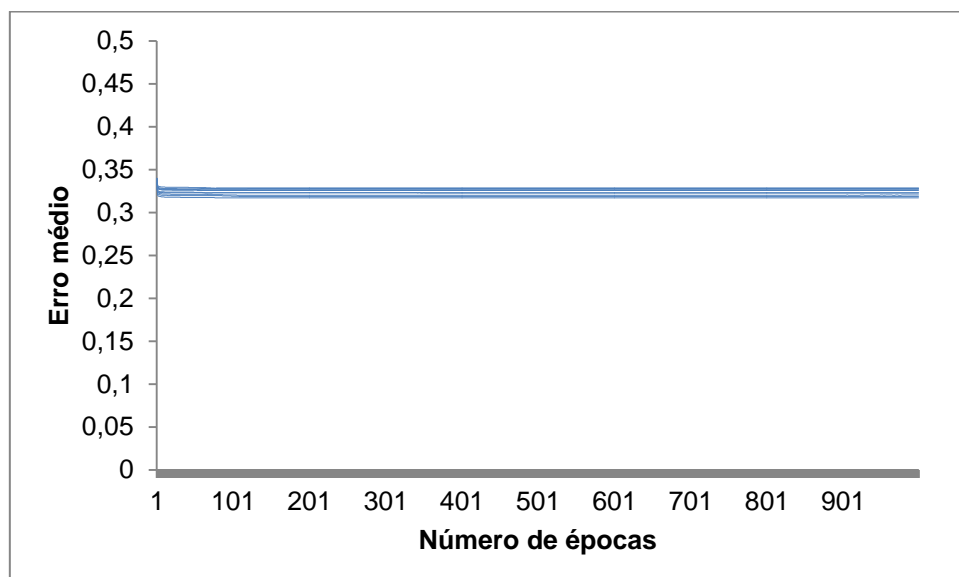


FIGURA 28- CONVERGÊNCIA DO ERRO DA REDE R2 NA SEXTA FASE DE TREINAMENTO
FONTE: A Autora, 2016.

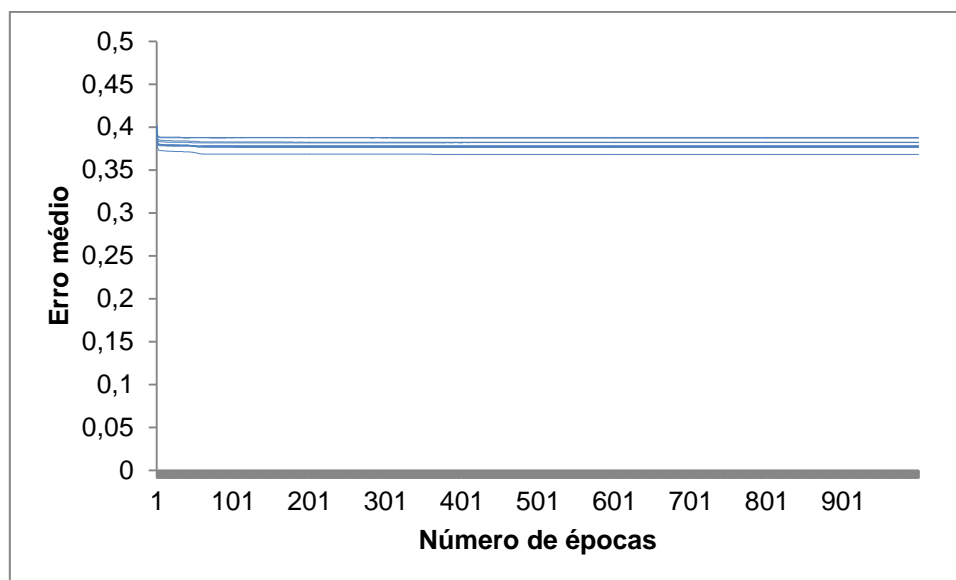


FIGURA 29- CONVERGÊNCIA DO ERRO DA REDE R3 NA SEXTA FASE DE TREINAMENTO
FONTE: A Autora, 2016.

Assim, chegou-se às RNAs ótimas para a previsão do fenômeno nevoeiro para uma, duas e três horas: as séries temporais apresentadas como dados de entrada são constituídas apenas de dados de visibilidade de quatro horas anteriores, no formato bipolar; a camada intermediária deve conter dois neurônios ocultos; e, para o treinamento, devem-se ajustar os dados da amostra de maneira que os casos de nevoeiro e não-nevoeiro se mantenham compatíveis, neste caso, como haviam, aproximadamente, 20 vezes mais dados de não-nevoeiro, usou-se razão de ajuste igual a 20. A Tabela 34, a seguir, apresenta a configuração final das três redes.

TABELA 34- MELHOR DESEMPENHO DA SEXTA ETAPA DE TREINAMENTO

RNA	Horas anteriores	Variáveis	Neurônios ocultos	Razão de Ajuste
R1	4	Vis	2	20
R2	4	Vis	2	20
R3	4	Vis	2	20

FONTE: A Autora, 2016.

4.3.2 Operação das Redes Neurais Artificiais ótimas

O programa de operação possibilita que as configurações ótimas, obtidas na fase de treinamento da RNA, sejam, finalmente, utilizadas na previsão do fenômeno de interesse. O programa de operação utiliza os

parâmetros de normalização dos dados e a configuração das posições dos centros, desvios padrões e pesos sinápticos das 10 redes obtidas no treinamento (validações cruzadas) para prever a ocorrência ou a não ocorrência do fenômeno. Neste caso, o programa de operação determina a previsão fornecida por cada uma das 10 RNAs, e como previsão final, informa a classificação predominante.

Para os dados do conjunto a ser submetido à operação, foram tomadas as mesmas informações de visibilidade disponíveis nas mensagens METAR utilizadas na previsão através do método de persistência apresentado na Seção 4.2, as quais não foram empregadas no treinamento e ajuste das RNAs.

Utilizando-se, então, os dados supracitados para a operacionalização das RNAs ótimas, obtiveram-se os resultados apresentados na Tabela 35.

TABELA 35- DESEMPENHO DA OPERAÇÃO DAS RNA

	PC	ISC	V	RFA	PD
R1	0,97	0,45	1,47	0,48	0,77
R2	0,97	0,33	1,47	0,59	0,61
R3	0,96	0,25	1,47	0,66	0,49

FONTE: A Autora, 2016.

Foi possível observar que, para R1, R2 e R3, as 10 RNAs empregadas na validação cruzada de cada rede forneceram previsões iguais: caso uma previsse nevoeiro, todas as demais também previam, e vice-versa. Este comportamento pode indicar um ótimo ajuste final das RNAs, onde as classes ficaram muito bem definidas, evitando que algum vetor apresentado pertencesse hora a uma classe, hora a outra.

Todos os treinamentos foram realizados com 1000 iterações, de modo a garantir o *overtraining*. No entanto, é possível observar nos gráficos de evolução do erro, que a convergência foi alcançada com aproximadamente 100 iterações. Além disso, a proximidade entre as linhas dos gráficos das diferentes redes, com diferentes grupos de treinamento e a pequena diferença apresentada entre os valores dos erros (conforme pode ser observado no Anexo B) sugere o alcance do mínimo global, ou mínimos locais muito próximos.

Assim como o programa de treinamento, o programa de operação também foi cedido pelo Professor Ricardo Almeida, e a linguagem computacional utilizada foi FORTRAN 95.

4.4 DISCUSSÃO DOS RESULTADOS

Dispondo todos os resultados dos experimentos de previsão obtidos através das técnicas de climatologia, persistência, e através do emprego de RNAs em uma única tabela, tal como a Tabela 36, a seguir, é possível comparar o desempenho destas três técnicas de previsão. Destaca-se aqui, que as previsões por persistência e através do emprego de RNA RBF foram realizadas utilizando-se os dados de visibilidade retirados do código METAR do ano de 2015; e as previsões por climatologia utilizaram o banco de registros do Sistema WEBMET, entre os anos de 1997 e 2014, pois algumas informações utilizadas nesta técnica, como a cobertura total de nuvens, não são disponibilizadas no código METAR.

TABELA 36- COMPARATIVO ENTRE OS RESULTADOS OBTIDOS

Horas de previsão	Técnica de previsão	PC	ISC	V	RFA	PD
+1h	Climatologia	0,95	0,04	6,46	0,96	0,28
	Persistência	0,99	0,61	1	0,24	0,6
	RNA	0,97	0,45	1,47	0,48	0,77
+2h	Climatologia	0,95	0,03	6,46	0,96	0,24
	Persistência	0,98	0,42	1	0,41	0,59
	RNA	0,97	0,33	1,47	0,59	0,61
+3h	Climatologia	0,95	0,03	6,46	0,97	0,21
	Persistência	0,97	0,32	1	0,52	0,48
	RNA	0,96	0,25	1,47	0,66	0,49

FONTE: A Autora, 2016.

É possível notar que a previsão do fenômeno através de climatologia resulta índices bastante inferiores aos índices obtidos através da técnica de persistência, e através do emprego da RNA. Estes últimos têm desempenhos

bastante semelhantes, sendo que, seguindo a hierarquia entre as métricas de avaliação de desempenho estabelecida na Seção 3.3, pode-se dizer que os resultados obtidos a partir do emprego da RNA foram, ainda, superiores aos resultados da persistência.

5 CONCLUSÕES

Analisando-se os desempenhos observados nos experimentos de previsão através das técnicas de climatologia, persistência e através do emprego de RNAs, podem-se destacar alguns resultados significativos das três técnicas.

Sobre a climatologia, pode-se concluir que esta técnica não produziu bons resultados para previsão do fenômeno nevoeiro. O método foi capaz de prever, aproximadamente, 20% dos eventos, e a razão de falso alarme foi de mais de 95%. Pode-se dizer que a climatologia é uma ferramenta bastante útil para a compreensão do comportamento das variáveis preditoras do fenômeno nevoeiro, tanto, que foi possível estabelecer um panorama destas variáveis quando na ocorrência do fenômeno. No entanto, assumir que, na ocorrência concomitante destas condições, haverá a formação do fenômeno, mostrou, através dos índices exibidos na Tabela 19, que a tendência é superestimar a ocorrência dos eventos.

Sobre a técnica de persistência, pode-se dizer que para o fenômeno nevoeiro, no Aeroporto Afonso Pena, esta técnica produz bons resultados, visto que o fenômeno em questão não sofre muitas alternâncias entre ocorrência e não-ocorrência. Após estabelecida a condição, esta permanece por várias horas sucessivas. Desta forma, a previsão por persistência é uma forma simples de projeção do fenômeno nevoeiro, para até três horas, com índices de acerto razoáveis.

Sobre a previsão através de RNA é possível notar que previsões de restrições de visibilidade por nevoeiro podem ser feitas, apenas, a partir de séries temporais de 4 horas de outros dados de visibilidade. Informações que, inicialmente, se imaginavam ser importantes, como direção e velocidade do vento, pressão, cobertura de nuvens, e umidade relativa não se mostraram significativas quando combinadas às informações de visibilidade. Este comportamento, talvez se explique pelo fato de que o valor do alcance visual horizontal é diretamente influenciado pelo comportamento de outras variáveis meteorológicas. Assim, fornecer dados de visibilidade juntamente às informações que a geraram, pode ser, para a RNA, uma redundância de

informações, e, a partir de informações repetidas a rede não é capaz de tomar novas decisões. Daí a estagnação no desempenho já na primeira combinação entre variáveis. Além disso, as forçantes sinóticas mostraram não afetar o desempenho da rede. Isso talvez se explique pelo mesmo motivo já citado: as forçantes sinóticas afetam as variáveis em superfície, que por sua vez afetam os valores de visibilidade. Por fim, todas as variáveis, de altitude ou superfície, acabam por se resumir na própria variável visibilidade.

Nota-se também, uma queda no desempenho apresentado entre o treinamento e a operação. Esse comportamento talvez se explique pelo tamanho do conjunto de dados utilizado, ou por uma possível atipicidade climática neste ano específico.

Comparado aos desempenhos verificados em outros trabalhos, é possível afirmar que os resultados alcançados nesta pesquisa não só podem auxiliar na previsão dos eventos de nevoeiro, como também contribuem ainda mais com as pesquisas que buscam, há anos, a maneira mais adequada de tratar esta problemática.

Como sugestão para próximas pesquisas, propõe-se que sejam testadas outras combinações entre as variáveis meteorológicas, com séries temporais maiores, e, simultaneamente, aplicar um método como a análise de componentes principais para otimizar a extração de informações pela RNA, pois pode haver alguma combinação específica, que permita à rede neural compreender de forma mais refinada o comportamento do fenômeno em questão; além de se estudar o emprego de uma única RNA para realizar as previsões para os três prazos de validade simultaneamente.

REFERÊNCIAS

ALMEIDA, L. A. **Análise de frequência da ocorrência de nevoeiro no Aeroporto Internacional Salgado Filho- Porto Alegre/RS**. Relatório de Pesquisa final do Curso de Especialização em Meteorologia Aeronáutica (MET 001) do Instituto de Controle do Espaço Aéreo (ICEA), 2008. Disponível em < http://www.redemet.aer.mil.br/uploads/2015/03/ocorr_nev_sbpa.pdf>. Acesso em 03/02/2016 às 16:34.

ALMEIDA, M. V. **Aplicação de técnicas de Redes Neurais Artificiais na previsão de curtíssimo prazo da visibilidade e teto para o Aeroporto de Guarulhos- SP**. Rio de Janeiro, 206 p. Tese (Doutorado em Engenharia Civil)- Faculdade de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

BALLARD, S. P.; GOLDING, B. W.; SMITH, R. N. B. "Mesoscale model experiment forecast of the Haar of northeast Scotland". American Meteorological Society: **Monthly Weather Review**, v. 119, n. 9, pp. 2107-2123, Sep. 1991.

BARRY, R. G.; CHORLEY, R. J. **Atmosphere, Weather and Climate**. 8th ed. London: Routledge, 2003.

BONEH, T. *et al.* "Fog Forecasting for Melbourne Airport Using a Bayesian Decision Network". American Meteorological Society: **Weather and Forecasting**, v. 30, n. 5, pp. 1218-1233, Oct. 2015.

BRASIL, Comando da Aeronáutica. Departamento de Controle do Espaço Aéreo. **Manual de Estações Meteorológicas de Superfície**. MCA 105-2. Rio de Janeiro, 2015.

BRASIL, Comando da Aeronáutica. Departamento de Controle do Espaço Aéreo. **Manual de Operação do WEBMET**. MCA 105-16. Rio de Janeiro, 2015.

CASTRO, F. C.; CASTRO, M. C. F. **Radial Basis Function**. Porto Alegre: PUCRS, 2002. Disponível em < http://www.feng.pucrs.br/~decastro/pdf/RNA_C5.pdf> Acesso em 23/07/2016 às 15:53.

FABBIAN, D.; DEAR, R; LELLYETT, S. "Application of Artificial Neural Network Forecasts to Predict Fog at Canberra International Airport". American Meteorological Society: **Weather and Forecasting**, v. 22, n. 2, pp. 372-381, Apr. 2007.

HANSEN, B. "A Fuzzy Logic-Based Analog Forecasting System for Ceiling and Visibility". American Meteorological Society: **Weather and Forecasting**, v. 22, n. 6, pp. 1319-1330, Dec. 2007.

INFRAERO. **Anuário Estatístico Operacional 2015**. Brasília, 2016. Disponível em <http://www.infraero.gov.br/index.php/br/estatisticas/estatisticas.html>. Acesso em 04/07/2016 às 18:32.

INFRAERO. **Aeroporto Internacional Afonso Pena**. Disponível em <http://www.infraero.gov.br/index.php/br/aeroportos/parana/aeroporto-afonso-pena.html>. Acesso em 07/08/2016 às 17:53.

KOZAR, M. E.; MISRA, V.; POWELL, M. D. "Hindcasts of Integrated Kinetic Energy in Atlantic Tropical Cyclones: A Neural Network Prediction Scheme". American Meteorological Society: **Monthly Weather Review**, v. 144, n. 12, pp. 4591-4603, Nov. 2016.

Mapa da cidade de São José dos Pinhais. Disponível em <https://www.google.com.br/maps?hl=pt-BR&tab=wl>. Acesso em 03/08/2016 às 09:13

Mapa de altitude do estado do Paraná. Disponível em <http://files.professoralexeinowatzki.webnode.com.br/200000388-4c3464d2e7/hipso%20pr.JPG>. Acesso em 29/09/2016 às 01:26.

Mapa do Brasil. Disponível em <http://www.mineropar.pr.gov.br/arquivos/Image/geoturismo/geologia/geologia01.jpg>. Acesso em 29/09/2016 às 02:11.

MARZBAN, C.; LEYTON, S.; COLMAN, B. "Ceiling and Visibility Forecasts via Neural Networks". American Meteorological Society: **Weather and Forecasting**, v. 22, n. 3, pp. 466-479, Jun. 2007.

MATHIEU, J. A.; AIRES, F. "Statistical Weather-Impact Models: An Application of Neural Networks and Mixed Effects for Corn Production over the United States". American Meteorological Society: **Journal of Applied Meteorology and Climatology**, v. 55, n. 11, pp. 2509-2527, Nov. 2016.

OLIVEIRA, L. L.; VIANELLO, R. L.; FERREIRA, N. J. **Meteorologia Fundamental**. 1 ed. Erechim: EdiFAPES, 2001.

PATIL, K.; DEO, M. C.; RAVICHANDRAN, M. "Prediction of Sea Surface Temperature by Combining Numerical and Neural Techniques". American Meteorological Society: **Journal of Atmospheric and Oceanic Technology**, v. 33, n. 8, pp. 1715-1726, Aug. 2016.

REDEMET. **METAR**. Disponível em www.redemet.aer.mil.br.

ROQUELAURE, S.; BERGOT, T. "Contributions from a Local Ensemble Prediction System (LEPS) for Improving Fog and Low Cloud Forecasts at Airports". American Meteorological Society: **Weather and Forecasting**, v. 24, n. 1, pp. 39-52, Feb. 2009.

ROSETTE, A. C. **Previsão a Curtíssimo Prazo de Eventos Meteorológicos Extremos para a área de influência do Aeroporto Internacional Tom Jobim utilizando Técnicas de Redes Neurais Artificiais**. Rio de Janeiro, 119 p. Dissertação (Mestrado em Meteorologia), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

Representação de um neurônio biológico. Disponível em <http://www.biologia.seed.pr.gov.br/modules/galeria/detalhe.php?foto=256&evento=3>. Acesso em 03/08/2016 às 08:13

SAHA, M.; MITRA, P. "Recurrent neural network based prediction of indian summer monsoon using global climatic predictors". **XXIX International Joint Conference on Neural Networks**, Vancouver, Canada, 2016.

SHANK, D.B.; HOOGENBOOM, G.; MCCLENDON, R.W. "Dewpoint Temperature Prediction Using Artificial Neural Networks". American Meteorological Society: **Journal of Applied Meteorology and Climatology**, v. 47, n. 6, pp. 1757-1769, Jun. 2008.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas**. 1 ed. São Paulo: Artliber, 2010.

SILVA, N. B. **Aplicação de Métodos Estatísticos e Redes Neurais no Pós-Processamento do Produtos de Previsão Numérica de Tempo**. Rio de Janeiro, 216 p. Dissertação (Mestrado em Engenharia Civil)- Faculdade de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

Universidade de Wyoming. **Sondagens**. Disponível em <http://weather.uwyo.edu/upperair/sounding.html>

VAREJÃO-SILVA, M. A. **Meteorologia e Climatologia**. 2 ed. Brasília: INMET. Pax, 2001.

WILKS, D. S. **Statistical Methods in the Atmospheric Sciences**; 2 ed. Cornell University: Department of Earth and Atmospheric Sciences. 2006.

Anexo A- Correlação linear entre as diversas variáveis e a variável visibilidade

Horas	vis	nuvem	precip	vel	temp	pressão	PO	UR	GPH500	GPH850	temp850	PO850	temp700	PO700	vel250
0	1	-0,3092	-0,1439	0,23607	-0,11	-0,0904	-0,0003	-0,4245	0,01017	0,00744	-0,0537	-0,0717	0,04887	-0,0448	0,00481
1	0,81173	-0,2759	-0,0696	0,24614	-0,1204	-0,0644	-0,0038	-0,42	0,00618	0,00747	-0,0443	-0,0722	0,0532	-0,0479	0,00262
2	0,65827	-0,2367	-0,0344	0,23858	-0,1149	-0,0529	-0,0044	-0,3934	0,00248	0,0075	-0,0337	-0,0711	0,05627	-0,051	-0,002
3	0,5321	-0,1994	0,00015	0,21803	-0,0993	-0,0534	-0,0038	-0,3537	-0,0012	0,00753	-0,0177	-0,0646	0,05586	-0,0527	-0,0056
4	0,42584	-0,1645	0,00526	0,1912	-0,0791	-0,0545	-0,0023	-0,3054	-0,0057	0,00757	-0,0049	-0,0536	0,05333	-0,0525	-0,0056
5	0,33779	-0,1324	0,02872	0,16123	-0,0558	-0,0622	-0,0004	-0,2517	-0,0094	0,0076	0,00256	-0,0436	0,0495	-0,051	-0,0057
6	0,26609	-0,1041	0,04691	0,12919	-0,0276	-0,0665	0,0023	-0,1961	-0,0129	0,00763	0,01067	-0,0337	0,04564	-0,0466	-0,0056
7	0,20707	-0,08	0,03865	0,09605	-0,0006	-0,076	0,00518	-0,1427	-0,0172	0,00689	0,01685	-0,0249	0,04167	-0,0406	-0,0056
8	0,15906	-0,0595	0,02066	0,06472	0,02517	-0,0762	0,0068	-0,091	-0,0196	0,00799	0,01133	-0,0171	0,03856	-0,0338	-0,0055
9	0,12199	-0,0423	0,00707	0,03657	0,05194	-0,0775	0,00682	-0,0425	-0,0222	0,00801	0,01402	-0,0076	0,03637	-0,0278	-0,0055
10	0,09399	-0,0297	0,0156	0,01051	0,06722	-0,0724	0,0057	0,00395	-0,0238	0,00801	0,0192	-0,0015	0,03616	-0,0255	-0,0023
11	0,07501	-0,0201	0,01684	-0,0112	0,07868	-0,0631	0,00367	0,04677	-0,0244	0,00646	0,02702	0,00083	0,03717	-0,0231	0,00183
12	0,06308	-0,0154	0,01883	-0,0317	0,08758	-0,0545	0,00065	0,0885	-0,0243	0,0049	0,03269	-0,0004	0,03805	-0,0222	0,00509

Anexo B- Evolução do erro para as Redes R1, R2 e R3 em cada fase de treinamento

R1		Fases de treinamento					
		1	2	3	4	5	6
Iterações	1	0,196376	0,196376	0,155637	0,155637	0,233024	0,23472
	2	0,196196	0,196196	0,146572	0,146572	0,211534	0,23559
	3	0,195627	0,195627	0,148313	0,148313	0,235349	0,234098
	4	0,211037	0,211037	0,149391	0,149391	0,245024	0,231337
	5	0,194361	0,194361	0,147152	0,147152	0,226291	0,232713
	6	0,19703	0,19703	0,154938	0,154938	0,238926	0,232835
	7	0,198134	0,198134	0,146834	0,146834	0,24536	0,230433
	8	0,198703	0,198703	0,150851	0,150851	0,223223	0,232047
	9	0,197462	0,197462	0,150376	0,150376	0,23957	0,230561
	10	0,19577	0,19577	0,148015	0,148015	0,239911	0,233107
Variação entre o erro máximo e mínimo (em %)		7,901761	7,901761	5,824579	5,824579	13,78605	2,188595

R2		Fases de treinamento					
		1	2	3	4	5	6
Iterações	1	0,190989	0,190989	0,195864	0,195864	0,322404	0,326354
	2	0,199268	0,199268	0,19042	0,19042	0,323064	0,327376
	3	0,196714	0,196714	0,198543	0,198543	0,320909	0,320106
	4	0,21055	0,21055	0,202967	0,202967	0,324773	0,327388
	5	0,202053	0,202053	0,213582	0,213582	0,32499	0,323024
	6	0,202702	0,202702	0,206634	0,206634	0,3257	0,317097
	7	0,188544	0,188544	0,211248	0,211248	0,325418	0,325505
	8	0,191035	0,191035	0,200295	0,200295	0,324732	0,327036
	9	0,193269	0,193269	0,192314	0,192314	0,325802	0,318621
	10	0,200577	0,200577	0,190293	0,190293	0,323742	0,322659
Variação entre o erro máximo e mínimo (em %)		10,45162	10,45162	10,9037	10,9037	1,501892	3,143217

R3		Fases de treinamento					
		1	2	3	4	5	6
Iterações	1	0,222749	0,222749	0,232215	0,232215	0,38549	0,376806
	2	0,223381	0,223381	0,225139	0,225139	0,371103	0,376401
	3	0,219032	0,219032	0,221061	0,221061	0,373385	0,387826
	4	0,225409	0,225409	0,218885	0,218885	0,370162	0,387235
	5	0,220336	0,220336	0,228136	0,228136	0,380458	0,377774
	6	0,21832	0,21832	0,225971	0,225971	0,365129	0,368062
	7	0,232409	0,232409	0,225331	0,225331	0,388033	0,377782
	8	0,227879	0,227879	0,22142	0,22142	0,380069	0,377268
	9	0,2172	0,2172	0,229436	0,229436	0,374397	0,382439
	10	0,212006	0,212006	0,234425	0,234425	0,387665	0,381631
Variação entre o erro máximo e mínimo (em %)		8,779096	8,779096	6,628616	6,628616	5,902639	5,096078